Titre: Title:	Multi-Criteria Inventory Classification and Root Cause Analysis Based on Logical Analysis of Data
Auteur: Author:	Yuchang Wu
Date:	2016
Type:	Mémoire ou thèse / Dissertation or Thesis
Référence: Citation:	Wu, Y. (2016). Multi-Criteria Inventory Classification and Root Cause Analysis Based on Logical Analysis of Data [Mémoire de maîtrise, École Polytechnique de Montréal]. PolyPublie. https://publications.polymtl.ca/2333/

URL de PolyPublie: PolyPublie URL:	https://publications.polymtl.ca/2333/
Directeurs de recherche: Advisors:	Soumaya Yacout
Programme: Program:	Maîtrise recherche en génie industriel

UNIVERSITÉ DE MONTRÉAL

MULTI-CRITERIA INVENTORY CLASSIFICATION AND ROOT CAUSE ANALYSIS BASED ON LOGICAL ANALYSIS OF DATA

YUCHANG WU

DÉPARTEMENT DE MATHÉMATIQUES ET DE GÉNIE INDUSTRIEL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

MÉMOIRE PRÉSENTÉ EN VUE DE L'OBTENTION

DU DIPLÔME DE MAÎTRISE ÈS SCIENCES APPLIQUÉES

(GÉNIE INDUSTRIEL)

OCTOBRE 2016

UNIVERSITÉ DE MONTRÉAL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

Ce mémoire intitulé :

MULTI-CRITERIA INVENTORY CLASSIFICATION AND ROOT CAUSE ANALYSIS BASED ON LOGICAL ANALYSIS OF DATA

présenté par : WU Yuchang

en vue de l'obtention du diplôme de : Maîtrise ès sciences appliquées

a été dûment accepté par le jury d'examen constitué de :

M. ADJENGUE Luc-Désiré, Ph. D., président

Mme YACOUT Soumaya, D. Sc., membre et directeur de recherche

M. OUALI Mohamed-Salah, Doctorat, membre

DEDICATION

I delicate this thesis to my family for their endless love and support, and friends who helped me in my difficulties and encouraged me to follow my dreams.

ACKNOWLEDGEMENTS

First and foremost, I would like to express my sincere gratitude to my supervisor Professor Soumaya Yacout for her continuous support of my study and related research, for her profound guidance, constructive advice, and immense knowledge. I appreciate her patience and valuable opinion during writing of this thesis. With her confidence in me and support, the thesis is becoming possible.

I also want to thank my committee members, Dr. Luc-Désiré Adjengue and Dr. Mohamed-Salah Ouali for serving as my jury committee. Their constructive comments help me refine my thesis.

My gratitude and appreciation goes as well to all those colleagues from our department for their help and advice.

My sincere thanks also go to Jeffrey Mo who encouraged me in any possible way. Without his precious support, it would not be possible to come here to finish my study.

Last but not the least, I would like to thank my family: my parents and my sister for supporting me spiritually throughout the journey of study and my life in general.

RÉSUMÉ

La gestion des stocks de pièces de rechange donne un avantage concurentiel vital dans de nombreuses industries, en passant par les entreprises à forte intensité capitalistique aux entreprises de service. En raison de la quantité élevée d'unités de gestion des stocks (UGS) distinctes, il est presque impossible de contrôler les stocks sur une base unitaire ou de porter la même attention à toutes les pièces. La gestion des stocks de pièces de rechange implique plusieurs intervenants soit les fabricants d'équipement d'origine (FEO), les distributeurs et les clients finaux, ce qui rend la gestion encore plus complexe. Des pièces de rechange critiques mal classées et les ruptures de stocks de pièces critiques ont des conséquences graves. Par conséquent il est essentiel de classifier les stocks de pièces de rechange dans des classes appropriées et d'employer des stratégies de contrôle conformes aux classes respectives. Une classification ABC et certaines techniques de contrôle des stocks sont souvent appliquées pour faciliter la gestion UGS.

La gestion des stocks de pièces de rechange a pour but de fournir des pièces de rechange au moment opportun. La classification des pièces de rechange dans des classes de priorité ou de criticité est le fondement même de la gestion à grande échelle d’un assortiment très varié de pièces. L'objectif de la classification est de classer systématiquement les pièces de rechange en différentes classes et ce en fonction de la similitude des pièces tout en considérant leurs caractéristiques exposées sous forme d'attributs. L'analyse ABC traditionnelle basée sur le principe de Pareto est l'une des techniques les plus couramment utilisées pour la classification. Elle se concentre exclusivement sur la valeur annuelle en dollar et néglige d'autres facteurs importants tels que la fiabilité, les délais et la criticité. Par conséquent l’approche multicritères de classification de l'inventaire (MCIC) est nécessaire afin de répondre à ces exigences.

Nous proposons une technique d'apprentissage machine automatique et l'analyse logique des données (LAD) pour la classification des stocks de pièces de rechange. Le but de cette étude est d'étendre la méthode classique de classification ABC en utilisant une approche MCIC. Profitant de la supériorité du LAD dans les modèles de transparence et de fiabilité, nous utilisons deux exemples numériques pour évaluer l'utilisation potentielle du LAD afin de détecter des contradictions dans la classification de l'inventaire et de la capacité sur MCIC.

Les deux expériences numériques ont démontré que LAD est non seulement capable de classer les stocks mais aussi de détecter et de corriger les observations contradictoires en combinant l'analyse des causes (RCA). La précision du test a été potentiellement amélioré, non seulement par l'utillisation du LAD, mais aussi par d'autres techniques de classification d'apprentissage machine automatique tels que : les réseaux de neurones (ANN), les machines à vecteurs de support (SVM), des k-plus proches voisins (KNN) et Naïve Bayes (NB). Enfin, nous procédons à une analyse statistique afin de confirmer l'amélioration significative de la précision du test pour les nouveaux jeux de données (corrections par LAD) en comparaison aux données d'origine. Ce qui s'avère vrai pour les cinq techniques de classification. Les résultats de l'analyse statistique montrent qu'il n'y a pas eu de différence significative dans la précision du test quant aux cinq techniques de classification utilisées, en comparant les données d'origine avec les nouveaux jeux de données des deux inventaires.

ABSTRACT

Spare parts inventory management plays a vital role in maintaining competitive advantages in many industries, from capital intensive companies to service networks. Due to the massive quantity of distinct Stock Keeping Units (SKUs), it is almost impossible to control inventory by individual item or pay the same attention to all items. Spare parts inventory management involves all parties, from Original Equipment Manufacturer (OEM), to distributors and end customers, which makes this management even more challenging. Wrongly classified critical spare parts and the unavailability of those critical items could have severe consequences. Therefore, it is crucial to classify inventory items into classes and employ appropriate control policies conforming to the respective classes. An ABC classification and certain inventory control techniques are often applied to facilitate SKU management.

Spare parts inventory management intends to provide the right spare parts at the right time. The classification of spare parts into priority or critical classes is the foundation for managing a large-scale and highly diverse assortment of parts. The purpose of classification is to consistently classify spare parts into different classes based on the similarity of items with respect to their characteristics, which are exhibited as attributes. The traditional ABC analysis, based on Pareto's Principle, is one of the most widely used techniques for classification, which concentrates exclusively on annual dollar usage and overlooks other important factors such as reliability, lead time, and criticality. Therefore, multi-criteria inventory classification (MCIC) methods are required to meet these demands.

We propose a pattern-based machine learning technique, the Logical Analysis of Data (LAD), for spare parts inventory classification. The purpose of this study is to expand the classical ABC classification method by using a MCIC approach. Benefiting from the superiority of LAD in pattern transparency and robustness, we use two numerical examples to investigate LAD’s potential usage for detecting inconsistencies in inventory classification and the capability on MCIC.

The two numerical experiments have demonstrated that LAD is not only capable of classifying inventory, but also for detecting and correcting inconsistent observations by combining it with the Root Cause Analysis (RCA) procedure. Test accuracy improves potentially not only with the LAD technique, but also with other major machine learning classification techniques, namely artificial

neural network (ANN), support vector machines (SVM), k-nearest neighbours (KNN) and Naïve Bayes (NB). Finally, we conduct a statistical analysis to confirm the significant improvement in test accuracy for new datasets (corrections by LAD) compared to original datasets. This is true for all five classification techniques. The results of statistical tests demonstrate that there is no significant difference in test accuracy in five machine learning techniques, either in the original or the new datasets of both inventories.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

AHP	Analytic Hierarchy Process
ANNs	Artificial Neural Networks
BPNs	backpropagation networks
CRS	Constant Return to Scale
DEA	Data Envelopment Analysis
DMU	Decision Making Unit
FCM	Fuzzy C-means
IDEA	Imprecise DEA
KNN	k-Nearest Neighbours
LAD	Logical Analysis of Data
MCIC	Multi-Criteria Inventory Classification
MDA	Multiple Discriminant Analysis
NA	Not Applicable
NB	Naïve Bayes
OEM	Original Equipment Manufacturer
OVA	One-Versus-All
OVO	One-Versus-One
RBF	Radial Basis Function
RCA	Root Cause Analysis
SKUs	Stock Keeping Units
SVM	Support Vector Machine
VRS	Variable Return to Scale

CHAPTER 1 INTRODUCTION

Spare parts inventory management plays a vital role in maintaining a competitive advantage in many industries, from capital intensive companies to service networks, such as railways, airlines, telecommunication (Boylan & Syntetos, 2010; Sarmah & Moharana, 2015; Stoll, Kopf, Schneider, & Lanza, 2015) etc. The purpose of classification is to consistently classify spare parts into different classes based on the similarity of items with respect to their characteristics. Wrongly classified critical spare parts and the unavailability of those critical items would have severe consequences (Sarmah & Moharana, 2015). Spare parts inventory management involves all parties, from Original Equipment Manufacturer (OEM) to distributors and end customers. Various forms of classifications have been widely performed in spare parts inventory management, inventory forecasting or production management (van Kampen, Akkerman, & van Donk, 2012).

The classification of spare parts into priority or criticality classes is the foundation for managing a large-scale and highly diverse assortment of parts (Rezaei & Dowlatshahi, 2010). Due to the massive quantity of distinct Stock Keeping Units (SKUs), it is almost impossible to control inventory by individual item or pay the same attention to all items (Babai, Ladhari, & Lajili, 2015). It is crucial to classify inventory items into classes and employ appropriate control policies conforming to these classes. Therefore, an ABC classification and certain inventory control techniques are often applied to facilitate SKU management (Fu, Lai, Miao, & Leung, 2015).

This study will propose using the Logical Analysis of Data (LAD), a pattern based machine learning technique, for spare parts inventory classification. The details of the LAD technique will be described further in Chapter 3, Methodology.

1.1 Statement of the problem

Spare parts inventory management intends to provide the right spare parts at the right time. The key problem is how to balance the cost of holding inventory and the risk of stock shortages (Kennedy, Wayne Patterson, & Fredendall, 2002). Either to maximize the profit from inventory sales or minimize the cost of inventory, we need to understand the characteristics of the machinery itself and then carry out the classification of inventories. Research on the classification of spare

parts helps to understand the nature of machinery, but we have not yet fully interpreted its attributes (van Kampen et al., 2012).

ABC analysis is one of the most widely used techniques for classification (Rezaei & Salimi, 2015; Stoll et al., 2015). The classical ABC classification is based on Pareto's Principle (Ramanathan, 2006) . Inventory is sorted by the total annual dollar usage, which is decided by unit price multiplied by annual usage rate. Though the ABC analysis is very popular for its ease of use, it concentrates exclusively on annual dollar usage and overlooks other important factors (Yu, 2011). The focus on this single criterion has resulted in taking no notice of other criteria, such as reliability, lead time, criticality, replicability, demand volume and inventory cost, which have been considered essential factors for inventory classification (Altay Guvenir & Erel, 1998; Ng, 2007; Ramanathan, 2006; Sarmah & Moharana, 2015; Stoll et al., 2015).

Multi-criteria classification methods are mostly divided into two categories, which are mathematical models and intelligence-based machine learning techniques. Mathematical models for inventory classification include analytic hierarchy process (AHP)(Lolli, Ishizaka, & Gamberini, 2014; Shamsaddini, Vesal, & Nawaser, 2015), data envelopment analysis (DEA) (Tavassoli, Faramarzi, & Saen, 2014) and fuzzy-rule-based approach (Sarmah & Moharana, 2015). On the other hand, machine learning techniques contain fuzzy c-means (FCM) clustering (Keskin & Ozkan, 2013), genetic algorithm (GA)(Altay Guvenir & Erel, 1998), artificial neural networks (ANNs) (Fariborz Y. Partovi & Anandarajan, 2002) etc. More details on classification methods will be discussed in Chapter 2, Literature Review. Despite their popularity, these methods either rely on certain assumptions about the importance of factors, or increase the complexity by recalculating classification with new inventory items. Furthermore, inconsistency in classification has been commonly found due to experts' biases or inaccurate recordings.

1.2 Objective

The purpose of this study is to expand the classical ABC classification method by using a multi-criteria inventory classification approach based on the machine learning technique. The Logical Analysis of Data (LAD), a pattern based classification method, is proposed for our experiment. LAD is a machine learning technique which is capable of extracting useful knowledge in the form

of interpretable patterns from a dataset. The superiority of LAD is in its patterns transparency and robustness. In addition, it does not rely on any statistical techniques.

By utilizing the advantages of LAD technique, this thesis attempts to achieve the following specific objectives:

1. Extend the traditional ABC classification with multi-criteria by LAD;
2. Investigate the potential use of LAD by detecting inconsistencies of inventory classification;
3. Provide evidence of the capability of LAD for classification by comparing other machine learning classification techniques.

1.3 Organization of the Thesis

Chapter 2 introduces the literature review on inventory classification. The methodology of LAD is found in Chapter 3. The process of data analysis is described in Chapter 4. Chapter 5 shows our experimental results that establish the capability of LAD on spare parts inventory classification, which is tested on numerical examples. The statistical analysis results are summarized in Chapter 6, including the conclusions we have drawn from our research. Chapter 7 suggests several ideas for related future work. Following these concluding chapters is the bibliography.

CHAPTER 2 LITERATURE REVIEW

For decades, there has been plenty of research on inventory control and operations management with regards to the classification of products. Although there are a few excellent general reviews of classification on spare parts inventory, each to some extent reflects the researchers' personal research interests and expertise. Due to the complexity of machine working conditions and the breadth of research, a truly comprehensive review is probably impossible, and certainly beyond the scope of this thesis. Instead, we focus on ABC classification and its extension schemes, which constitute the major goal of the research. The following brief review presents the characteristics of classical ABC classification and mathematical model based methods in addition to the principles of the popular machine learning techniques for ABC classification.

2.1 Traditional ABC classification

The traditional ABC classification is based on Pareto's principle, also known as the 80-20 rule, which was developed at General Electric during the 1950s (Altay Guvenir & Erel, 1998; Keskin & Ozkan, 2013). The aim of ABC analysis is to categorize inventory into three classes, namely A (very important); B (moderately important) and C (relatively unimportant) (Hatefi, Torabi, & Bagheri, 2014). Class A includes all items within the cumulative value of 70-80%, class B includes all items with the cumulative value up to 95%, and the rest of the items are class Cs(Ng, 2007). Accordingly, each class is assigned a control level and a service level that are applied to all Stock Keeping Units (SKUs) in a specific class. More details on inventory control policies can be found in Silver, Pyke, and Peterson (1998).

ABC analysis is the most popular method for inventory classification by virtue of its clarity and capability. The classification of spare parts inventory is mostly based on the managerial efficiency concern and concentrates on the most valuable items (Braglia, Grassi, & Montanari, 2004). This practice inevitably overlooks other attributes of spare parts, such as lead time and reliability, and hardly satisfies the operations' requirements for high availability at a low cost. Some researchers have introduced second criterion criticality of spare parts, aside from annual dollar usage, to extend the ABC analysis (Duchessi, Tayi, & Levy, 1988). This approach actually involves several other parameters, such as lead time and expected failure, to determine the criticality. However, either one-dimensional or two-dimensional classification schemes have limitations on the separation of

important factors from all potential useful parameters. A number of researchers have proposed the use of multiple criteria, such as lead time, reliability and obsolescence, to extend the ABC classification. The next section provides a brief review of Multi-Criteria Inventory Classification (MCIC).

2.2 Multi-criteria inventory classification

2.2.1 Analytic hierarchy process

Although classical ABC analysis is best known for its easy implementation and simplicity, it has been criticized for solely focusing on dollar usage and overlooking other crucial factors for inventory classification. Since then, many multi-criteria classification methods have been developed. One of the more popular techniques adopted in inventory classification is the analytic hierarchy process (AHP). The AHP methodology was proposed for spare parts classification from a number of authors (Cebi, Kahraman, & Bolat, 2010;Gajpal, Ganesh, & Rajendran, 1994; F. Y. Partovi & Hopton, 1994). The AHP is a decision making tool for analyzing complex problems which involve multiple criteria. The theory of tree structured AHP technique is formulated as pairwise comparisons to facilitate the decision-support procedure, which starts by calculating the relative weight of each criterion at each layer of the hierarchy and assessing the overall evaluation of all alternatives at the base level of the hierarchy.

The AHP techniques of ABC classification have been widely used. The main difference among those AHP techniques is the adoption of diverse criteria in the evaluation process. For example, F. Y. Partovi & Burton (1993) used four attributes of spare parts, which are unit cost, procurement cost, demand range and lead time, to classify inventory. Gajpal Ganesh & Rajendran (1994) proposed a scheme with three criteria, status of availability, type of spares and lead time, to estimate the criticality of spare parts by using the AHP technique.

One of the advantages of the AHP methodology is its adaptability in combining with other advanced techniques such as Artificial Neural Networks (ANNs), fuzzy logic and Data Envelopment Analysis (DEA) (Hadi-Vencheh & Mohamadghasemi, 2011; Kabir & Hasin, 2013; Shamsaddini et al., 2015). This characteristic allows users to obtain benefits from other methods and achieve a better solution; however, the AHP technique requires personal knowledge to assign

the weight of criteria. This assumption may bring about the inconsistency of inventory classification.

2.2.2 Data envelopment analysis

Another dynamic technique is Data Envelopment Analysis (DEA), which was originally developed by Charnes, Cooper, and Rhodes (1978). The principle of the DEA is to measure the relative performance of each Decision Making Unit (DMU) with multiple inputs and outputs. Initially it was used to evaluate the efficiency of non-profit and public organizations. Since then, many researchers have extended the DEA models and have successfully applied the models to many fields, such as financial efficiency and environment performance and classification (Liu, Lu, Lu, & Lin, 2013). One of these useful applications is inventory classification.

Ramanathan (2006) developed a DEA-like model combined with weighted linear programming for inventory classification. The model transfers all criteria into scalar scores and yields optimal scores for each inventory item by using weighted linear optimization, and then classifies items into classes based on the score value of items. This method may take a very long time when encountering thousands of inventory items, which is very common in industries. Ng (2007) improved the formulation with an alternative weight linear programming to solve the time cost problem, but the step for ranking criteria completely depends on users' expertise and experience. This situation may lead to human bias and an inconsistency in inventory classification.

Most recently, researchers have proposed hybrid methods of DEA with other techniques, such as neural networks, fuzzy AHP and discriminant analysis (Hadi-Vencheh & Mohamadghasemi, 2011; Pendharkar, 2010; Tavassoli et al., 2014). The main difference among them is the way of calculating weights of the criteria of inventory items. Torabi, Hatefi, & Saleck Pay (2012) argued that most existing DEA models can only handle quantitative criteria. They developed a modified DEA-like model that takes both quantitative and qualitative criteria into consideration. The principle of the DEA-like model is that it applies concepts from an imprecise DEA (IDEA) model to ABC inventory classification.

2.3 Machine learning classification

With the development of computer science and artificial intelligence, machine learning techniques have been widely studied in many fields. One of the more common applications in machine learning is inventory classification.

2.3.1 Artificial neural networks

Artificial Neural Networks (ANN) is one of most popular techniques in machine learning. For instance, Partovi and Anandarajan (2002) introduced an Artificial Neural Networks (ANN) technique combining backpropagation and genetic algorithms into two learning methods for inventory classification purposes. The results are compared between the two learning methods and nonlinear relations among the criteria are discovered. But the meta-heuristics approach may be too difficult for inventory managers to understand and may result in less applicability in industry practices.

Simunovic, Simunovic, & Saric (2009) presented a model of neural networks to classify inventory items. They developed feed-forward neural networks trained by backpropagation and used minimum root mean square error for evaluating the performance. The final results show that the neural networks technique has better performance compared to the AHP method. Kabir & Hasin (2013) proposed an integration model of fuzzy AHP and neural networks for multi-criteria inventory classification. They adopted fuzzy AHP method to measure the weights of criteria of inventory items and similarly applied backpropagation to train the feed-forward neural networks. The performance of the model is assessed by the minimum mean absolute percentage of error between computed and predicted values.

2.3.2 Support vector machines

The support vector machine (SVM) is a supervised learning algorithm introduced by Vapnik (1995). The SVM classification method is based on the structural risk minimization principle (Yu, 2011). SVM approaches have been established as a popular machine learning tool in classification and aggression fields. Many applications of SVMs have been studied, including faults diagnosis, text classification, image detection, etc. (Guosheng & Guohong, 2008). Su, Zhou, & Mo, (2010) proposed a new classification scheme which is based on SVM to categorize the spare parts

inventory class. They employed the risk level as the indicator of inventory class according to the attributes of spare parts, such as importance, standardization level and replicability.

A comparison work was done between artificial-intelligence (AI)-based classification techniques and traditional multiple discriminant analysis (MDA) by Yu (2011). The AI-based techniques include support vector machines (SVMs), backpropagation neural networks (BPNs), and the k-nearest neighbour (k-NN). The results show the AI-based classification techniques have superiority over MDA.

2.3.3 K-nearest neighbours

The k-Nearest-Neighbours (k-NN) is another popular non-parametric classification and pattern recognition technique. The method is simple but effective in many cases (Gongde, Hui, Bell, Yaxin, & Greer, 2003). The principle of k-NN is to assign a new instance to the same class by determining the classification of instances that is closest to the new one. Selecting the k value is essential for a k-NN technique. An improved k-NN classification algorithm is proposed by Gong & Liu (2011). They developed a model that can dynamically get the value of k.

The k-NN classifier is an instance-based learning technique that requires computing the distance and ranking all training instances at each prediction, which is computationally expensive when classifying a lot of new instances or instances with many attributes. Another limitation is that the k-NN algorithm cannot learn anything from the training process and is not robust enough for noisy data (Bramer, 2013). To overcome the disadvantages, several researchers proposed modified models integrating other techniques. For example, Kalaivani & Shunmuganathan (2014) developed a k-NN classifier using a genetic algorithm to increase the capability by choosing appropriate attributes and achieving lower computational cost. Mejdoub & Ben Amar (2013) presented a scheme of k-NN algorithm to reduce attribute space by using the hierarchical classification technique.

2.3.4 Naïve Bayes

The Naïve Bayes (NB) model is a simple probabilistic classifier that is based on Bayes' theorem with the assumption of independence among any feature (Agarwal, Jain, & Dholay, 2015). The NB algorithm incorporates the prior probability and conditional probabilities into one formula for

estimating the probability of every possible classification (Bramer, 2013). The limitation of NB is the independence assumptions between attributes. As reviewed by Jiang, Wang, Cai, & Yan (2007), many researchers have tried to overcome the limitation. For example, Ratanamahatana & Gunopulos (2003) proposed a combining decision tree NB algorithm that can choose the most relevant attributes of the training set to improve the classification accuracy. Webb, Boughton, & Wang (2005) presented an approach to relax the attribute independence assumption by averaging all of a constrained class of classifiers (called one-dependence classifiers).

CHAPTER 3 METHODOLOGY

3.1 Introduction

In general, data mining or machine learning provides data-driven analysis methods to extract useful knowledge or patterns from datasets. Data mining or machine learning is often divided into two main categories: supervised learning (predictive) and unsupervised learning (descriptive) methods. Supervised learning is to predict outputs for unseen observations by learning a set of input-output pairs, such as classification and regression, etc. Unsupervised learning is to find human-interpretable patterns by only given inputs, such as clustering and deep learning (Murphy, 2012).

Classification is to build a model for predicting the class of unknown observations as accurately as possible by providing a set of labelled datasets, which establishes the purpose of our study. A large number of studies on classification try to solve two-class (binary) problems where a classifier is built to discriminate new observations from two classes. But in many situations, more than two classes are involved in classification problems, such as inventory classification (Hadi-Vencheh & Mohamadghasemi, 2011; Shamsaddini et al., 2015), image recognition (Foody & Mathur, 2004; Joshi, Porikli, & Papanikolopoulos, 2012), cancer classification (Rifkin et al., 2003; Rui, Anagnostopoulos, & Wunsch, 2007; Zainuddin & Ong, 2011), handwritten interpretation (F. Chang, Chou, Lin, & Chen, 2004; Ou, Murphey, & Lee, 2004; Srihari, 2000), text categorization (Lewis, Yang, Rose, & Li, 2004; Weizhu, Jun, Benyu, Zheng, & Qiang, 2007) and speech recognition (Nakamura et al., 2006; Wang, Wang, Lin, Jian, & Kuok, 2006; Yang et al., 2012).

The two most common used approaches for multiclass classification are One-Versus-All (OVA) (or One-vs-Rest) and One-Versus-One (OVO) (also called all-pairs or All-vs-All) schemes. The main idea is to decompose the multiclass problems into multiple two-class problems. Given an N-class dataset, the OVA method is used to build N different binary classifiers by using one of the techniques, such as LAD, SVM or NB, etc. For the *i*th classifier, let the positive observations be all the points in class *i*, and let the negative observations be all the points not in class *i*. Let f_i be the *i*th classifier. The new observation x is classified by

$$f\left(x\right) = \underset{i}{\operatorname{argmax}}\, f_i(x)$$

The OVO method is to build N (N-1) / 2 classifiers, one classifier to distinguish each pair of classes *i* and *j*. Let f_{ij} be the classifier where class *i* are positive observations and class *j* are negative. Please note $f_{ji} = -f_{ij}$. So the new observation x is classified by

$$f(x) = \underset{i}{\operatorname{argmax}} \left(\sum_{j} f_{ij}(x) \right)$$

Both OVA and OVO schemes are very simple and they were invented independently by many researchers. The choice between OVA and OVO methods is largely computational. It is more important to tune proper regularization classifiers as the underlying binary classifiers than to choose between OVA and OVO. For an overview study on OVA and OVO schemes, please refer to the article by Galar, Fernandez, Barrenechea, Bustince, & Herrera (2011). A good comparison between OVA and OVO schemes can be found in the article Duan, Rajapakse, & Nguyen (2007).

In this thesis, we adopt the OVA scheme to build LAD classifiers to solve multiclass classification problems; specifically, inventory classification.

3.2 Logical analysis of data

The Logical Analysis of Data (LAD) is a relatively new technique that intends to detect structural information and extract favorable knowledge in the style of interpretable patterns from datasets (Boros et al., 2000). This pattern-based supervised learning approach was initially presented by Crama, Hammer, & Ibaraki (1988). The Peter L. Hammer team plays a vital role in theoretical and applied developments of LAD. One of their successful applications of LAD was implemented in the medical field.

Recently the LAD technique has been studied with diverse applications, such as classification, feature selection, decision support, etc. (Boros et al., 2000). The advantage of LAD has enabled this technique to achieve plenty of applications in medical diagnosis, politics, economics, etc. A good review of LAD is presented by Alexe et al. (2007). A number of applications of LAD have been presented (S. Alexe et al., 2003; Dupuis, Gamache, & Pagé, 2012; P. L. Hammer & Bonates, 2006; Lejeune & Margot, 2011). The Yacout team was the first to apply the LAD technique to solve engineering problems, where most applications are in condition-based maintenance fields (Salamanca, 2007), such as equipment useful life prediction (Ragab, Ouali, Yacout, & Osman,

2014), rogue components detection (Mortada, Carroll Iii, Yacout, & Lakis, 2012), and fault diagnosis (Mortada, Yacout, & Lakis, 2013).

The LAD algorithm combines the theories of optimization, combinatorics and Boolean functions. Patterns, the essence of LAD technique's decision rules, are needed to define discriminant function to separate observations between positive and negative in a dataset. Several studies of multiclass LAD approaches are proposed (Avila-Herrera & Subasi, 2015; Moreira, 2000; Mortada et al., 2013). Mortada et al. (2013) presented an OVO style multiclass LAD algorithm using mixed integer linear programming (MILP) approach to pattern generation, which is inspired by Moreira (2000) and Ryoo & Jang (2009). Avila-Herrera & Subasi (2015) proposed an OVA style multiclass LAD algorithm which also uses the MILP approach for pattern generation. The OVA style multiclass LAD model is adopted in our multi-criteria inventory classification study.

The implementation of the LAD algorithm is divided into three steps: data binarization, pattern generation and theory formation, test and classification.

3.2.1 Data binarization

Before conducting an analysis, data must be binarized so that it can be readable by computers. Each observation is considered a vector of *m* attributes that usually are shown as the non-binary format. The binarization process is used to transform attributes into Boolean variable vectors of *n* binary attributes. The non-binary attributes can be sorted into two categories: nominal indicators (e.g. color and critical level) and numerical indicators (e.g. price). The binarization of such nominal attributes is accomplished in an easy way by transforming each value v_s of the attribute x into a Boolean variable $b(x,v_s)$ such that we obtain the formula below.

$$b(x, v_s) = \begin{cases} 1, & x = v_s \\ 0, & otherwise \end{cases}$$

As for numerical attributes, a practical binarization technique is to arrange attributes into a value-order style. Then a "cut-point", which is a level variable (or interval variable) to indicate the attribute belong to certain level, is introduced. It means that for each attribute *x* and cut-point *t (or cut-points t′, t″)* a Boolean variable *b(x, t)* (or *b(x, t′, t″)*) shall be introduced as the following:

$$b(x,t) = \begin{cases} 0, & x < t \\ 1, & x \geq t \end{cases}$$

$$\text{or } b(x,t',t'') = \begin{cases} 1, & t' < x < t'' \\ 0, & otherwise \end{cases}$$

The cut-point t is the average value of v_n and v_{n-1} ($v_n < v_{n-1}$) such that v_n marks as one (1) and v_{n-1} marks as zero (0) or vice versa. The interval cut-point is formed by every two cut-points.

One example binarization procedure is presented. Suppose we have a sample dataset shown in Table 3.1, where the first and third attributes are numerical, while the second one is nominal.

Table 3.1: Sample of dataset

Label	Obs.	x1	x2	x3
D+	1	1038	critical	7
	2	855	critical	3
	3	594	important	6
D-	4	455	important	4
	5	268	regular	7
	6	703	important	4

First we arrange numerical attributes from small to large so that we can calculate cut-points. Interval cut-points are also formed based on cut-points. So we easily obtain level and interval variables shown in Table 3.2 and Table 3.3. Once we have level and interval variables, the attributes can be transformed to the Boolean form shown in Table 3.4. Thus, the binarization of the dataset is done.

Table 3.2: Level variables of attributes

b1	b2	b3	b4	b5	b6	b7	b8	b9
x1≥524.5	x1≥648.5	x1≥779	x2=critical	x2=important	x2=regular	x3≥3.5	x3≥5	x3≥6.5

Table 3.3: Interval variables of attributes

b10	b11	b12	b13	b14	b15
524.5≤x1<648.5	524.5≤x1<779	648.5.5≤x1<779	3.5≤x3<5	3.5≤x3<6.5	5≤x3<6.5

Table 3.4: Binary of attributes

Obs.	b1	b2	b3	b4	b5	b6	b7	b8	b9	b10	b11	b12	b13	b14	b15
1	1	1	1	1	0	0	1	1	1	0	0	0	0	0	0
2	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0
3	1	0	0	0	1	0	1	1	0	1	1	0	0	1	1
4	0	0	0	0	1	0	1	0	0	0	0	0	1	1	0
5	0	0	0	0	0	1	1	1	1	0	0	0	0	0	0
6	1	1	0	0	1	0	1	0	0	0	1	1	1	1	0

3.2.2 Pattern generation and theory formation

Patterns play a vital role in LAD algorithm due to offering straightforward interpretation of datasets. The form of the patterns is the combination of attributes to formalize rules that can define homogenous subsets in the dataset. Once attributes are binarized, a pattern can be represented as a Boolean term of a conjunction of literals:

$$P = \bigwedge_{i \in Kp} X_i;\ K_p \subseteq \{1,2,\cdots,n\}$$

Where n is the number of attributes in datasets, *Kp* is the set of attributes in pattern P. X_i is a binary variable and $\overline{X_i}$ is its negation. The number of literals is called the degree of a pattern. A pure positive (negative) pattern is defined as its attributes covering at least one positive (negative) observation but not any negative (positive) observation. One of the most common ways of pattern generation is the combinatorial enumeration technique, such as a top-down or a bottom-up approach. The top-down approach starts by regarding all uncovered observations as patterns and removes literals one by one for those patterns until achieving a prime pattern. The bottom-up approach begins with a term of degree one which covers some positive observations. If the term only covers positive observations but not any negative ones, it is a pattern. Otherwise, literals are added to the term one by one until reaching a pattern. The details of this approach can be found in the article by Boros et al. (2000). The enumerative technique to pattern generation is a time-consuming task. For terms of degree *d* with *n* Boolean variables, the number of candidate patterns can grow to $2^d \binom{n}{d}$. A number of studies on types of patterns and pattern generation methods have

been proposed for thrifty patterns (G. Alexe et al., 2007; G. Alexe & Hammer, 2006; Boros et al., 2000; Peter L. Hammer, Kogan, Simeone, & Szedmak, 2004; Ryoo & Jang, 2009).

In our study we adopt the MILP approach to generate patterns (Avila-Herrera & Subasi, 2015; Mortada et al., 2013; Ryoo & Jang, 2009). Given an ***N***-class binary dataset $\mathbb{Z}$ with *m* observations and *n* attributes, in such way we have C_i *(i =1, 2, ⋯, N)* standing for the corresponding classes. Let Pc_p be a pattern covering some observations (coverage denoted as Cov(Pc_p)) from class C_p and none of observations from class C_k, $(k \neq p)$.

The variables involved in the pattern generation algorithm are the pattern degree *d*, the Boolean pattern vector ***y*** defining the composition of the pattern found, and the coverage vector ***w***.

Constraints should be satisfied to generate a pattern Pc_p for the objective function:

(1) The Boolean vector $\boldsymbol{y} = (y_1, y_2, \cdots, y_{2n}) \in \{0, 1\}^{2n}$ has such elements of binarized training dataset that if $y_j = 1$ for some $j = 1,2, \cdots, n$, then the literal x_j (associated with the *j-th* attribute in the dataset $\mathbb{Z}$) is included in pattern Pc_p. Similarly, if $y_{n+j} = 1$ then literal $\overline{x}_j$ (complementary element of x_j) is included in pattern Pc_p. So, each binarized attribute y_j in the training dataset can be expressed by a literal x_j or its complement $\overline{x}_j$ in a pattern. Because a pattern cannot include both the literal x_j and $\overline{x}_j$, and the degree *d* of a pattern is associated with the number of literals, we have the constraints below:

$$y_j + y_{n+j} \leq 1, j = 1, 2, \cdots, n. \tag{3.1}$$

$$\sum_{j=1}^{2n} y_j = d, 1 \leq d \leq n. \tag{3.2}$$

(2) A binary vector $\boldsymbol{w} = (w_1, w_2, \cdots, w_m)$ ($w_i \in \{0, 1\}$, $1 \leq i \leq m$) is defined so that its elements are associated with the coverage of pattern Pc_i. The elements w_i of vector ***w*** are the variables to minimize in the set covering problem so that for $1 \leq i \leq m$, w_i equals to one if observation o_i from class *Cp* is not covered by pattern Pc_p, otherwise equals to zero.

(3) Build an augmented matrix $M = \left[\mathbb{Z} \mid \overline{\mathbb{Z}}\right]$, where $\overline{\mathbb{Z}}$ is acquired from $\mathbb{Z}$ by switching zero elements to one and one elements to zero. Let vector $\mathbf{u} = M\mathbf{y}$. The generated pattern must be able to cover at lease one observation *Oi* from class C_p ($o_I \in C_p$), however it is not

required to cover all the observations in class C_p. Hence generating a pure pattern Pc_p with degree d has the constraints:

$$u_i + nw_i \geq d, i \in I_p; \quad (3.3)$$

A pattern for class k should not cover any observations from class p $(k \neq p)$. So the dot product of vector u_i $(i \in I_k)$ must be less than the degree d of p:

$$u_i \leq d - 1, i \in I_k, k = 1, \cdots, K; k \neq p; \quad (3.4)$$

The formation of the pattern can be inferred from vector $\boldsymbol{y}$, and its coverage vector $\boldsymbol{w}$. Therefore, a pure pattern Pc_p connected with class C_p, $(1 \leq p \leq K)$ is determined by solving the optimal solution of the MILP problem below (Avila-Herrera & Subasi, 2015).

$$\begin{aligned} &Minimize \quad d + \sum_{i \in I_p} w_i \quad (3.5) \\ &s.t. \qquad (3.1), (3.2), (3.3), (3.4) \end{aligned}$$

A pattern P is a strong pattern if and only if there is no pattern P′ such that Cov(P′) ⊃ Cov(P). A pattern is a prime pattern if removal of any of its literals makes it a non-pattern. Avila-Herrera & Subasi (2015) proved that an optimal solution (u, y, w, d) of problem (3.5) can be formed a pattern with maximum coverage and minimum degree, which produces a strong prime pattern. The strong prime pattern Pc_p has the following form.

$$Pc_p = \bigwedge_{s1} X_j \bigwedge_{s2} \overline{X_j}; S1 = \{j : y_j = 1\}, S2 = \{j : y_{n+j} = 1\}, j = 1, 2, \cdots, n.$$

3.2.3 Test and classification

Suppose we have an N-class dataset $\mathbb{Z} = \mathbb{Z}_1 \cup \mathbb{Z}_2 \cup \ldots \cup \mathbb{Z}_N$ for training and testing where $\mathbb{Z}_k$ is the set of observations from class k $(k = 1, 2, \ldots, N)$.The corresponding multi-class LAD models are denoted as $\mathbb{S} = \mathbb{S}_1 \cup \mathbb{S}_2 \cup \ldots \cup \mathbb{S}_N$, ($\mathbb{S}_i \cap \mathbb{S}_j = \phi$; $i,j = 1, 2, \ldots, N$ and $i \neq j$). For each new observation ($O \notin \mathbb{Z}$), a score is calculated based on generated patterns which cover this new observation. The judgement task is solved by a discriminant function that produces a score for each class based on patterns covered that new observation. The class along with the highest score is the estimation class

for this new observation. The discriminant function for classification of new observation ($O \notin \mathbb{Z}$) is formulated as below.

$$\Delta(O) = \underset{n}{\operatorname{argmax}} \sum_{C_p \in S_n} w_{C_p} P_{C_p}(O),\ n = 1, \cdots, N$$

Patterns $P_{C_p} \notin \mathbb{S}_n$ (n=1, 2,…,N) in a support set have weights W_{C_p} associated coverage which is the ratio of the number of covered observations and all observations in that class C_P.

$$W_{C_p} = \frac{1}{|\mathbb{Z}_p|} \sum_{i \in I_{C_p}} P_{C_p}(O_i),\ \mathbb{Z}_p \subset \mathbb{Z},\ I_{C_p} = \left\{ i : o_i \in \mathbb{Z}_p, 1 \le p \le N \right\}$$

The class of observation O is estimated by the highest-class value of the discriminant function $\Delta(O)$ calculated. Test accuracy is evaluated by the most commonly used method of cross-validation. The testing stage is for examination of the multi-class process and inconsistency in the dataset. The process of data analysis based on the LAD test result is presented in next Chapter. Once the test accuracy reaches an acceptable level (normally above 90%), the multi-class LAD model is ready to make classification for new (unseen) observations by using the discriminant function.

CHAPTER 4 THE PROCESS OF DATA ANALYSIS

In this chapter, we introduce our data analysis process when applying the Logical Analysis of Data (LAD) machine learning technique. The process contains two steps, namely a classification analysis of LAD and Root Cause Analysis (RCA) for misclassified items. We run both steps reiteratively until all inconsistent items are corrected and an acceptable test accuracy is received.

4.1 Tools

The random partition of the training datasets and the testing datasets is done by using Matlab. The implementation of LAD analysis is done by cbmLAD software which is written in C++ Programming language at École Polytechnique de Montréal. Several articles use cbmLAD as analysis tool to implement the LAD technique (Bennane & Yacout, 2012; Mortada et al., 2012; Mortada & Yacout, 2011).

4.2 LAD classification analysis procedure

The procedure of LAD classification analysis is to explore the dataset and to detect any problems with the dataset.

1. Suppose we have an inventory dataset $\mathbb{Z} = \mathbb{Z}_A \cup \mathbb{Z}_B \cup \mathbb{Z}_C$ with three classes A, B and C. Firstly, we split the dataset $\mathbb{Z}$ into two disjoint datasets called training dataset $\mathbb{Z}_{TR}$ and testing dataset $\mathbb{Z}_{TS}$. So that we have the equations:

   $$\mathbb{Z} = \mathbb{Z}_{TR} \cup \mathbb{Z}_{TS}; \ \mathbb{Z}_{TR} \cap \mathbb{Z}_{TS} = \emptyset.$$

   The observations in dataset $\mathbb{Z}$ are set randomly into the training dataset $\mathbb{Z}_{TR}$ and the testing dataset $\mathbb{Z}_{TS}$, such as the proportions of the three classes in those subsets are the same as in the original dataset.

2. We execute cbmLAD on the training datasets and obtain the multi-class LAD classification and the corresponding patterns. Here we use 80% of dataset as the training dataset and the rest of the dataset as the testing dataset.

3. We execute cbmLAD on the testing datasets and calculate the test accuracy. The classification procedure by LAD is repeated on different training and testing sets either twenty or ten times. The number of repetition is chosen arbitrarily to ensure more persistent test results.

The test of twenty times is adopted in the first numerical example and ten times is for the second numerical example. The training datasets and testing datasets are randomly partitioned each time. Once the twenty (for first numerical example) or ten (for second numerical example) tests are finished, an average test accuracy is obtained by averaging the twenty tests' accuracies. Those twenty or ten tests are defined as one round of analysis.

4.3 Root cause analysis for misclassification

One of the most important advantages of LAD is the transparency, which means the generated patterns can be easily interpreted. We make use of this characteristic of LAD to investigate misclassifications. In order to understand misclassifications, we examine the attributes of observations and find out the reason for a misclassification by using a procedure called Root Cause Analysis (RCA).

RCA begins by checking for any contradictions or repetitions. Contradictions in this thesis refer to two or more observations with exactly same attributes' values but labelled as different classes. Since the basic assumption is that the attributes are enough to discriminate observations of different classes, all contradictory observations will be eliminated. Repetitions mean two or more observations are exactly the same. Since our objective in the training phase is to find patterns and the removal of repeated observation does not affect the pattern generation, under the condition of repetition, observations are considered as only one observation and other(s) are removed from dataset. The cbmLAD software is capable of detecting any contradictions or repetitions during the training and test phase.

Secondly, we take advantage of the logical interpretation of patterns by LAD. Patterns are generated during the training stage to reveal the characteristics of each class. Each pattern illustrates its coverage and shows corresponding weight of each pattern in the same class. The weight is the ratio of the number of covered observations by that pattern and all observations in the same class. The more observations covered by the pattern, the more weight the pattern has. In other

words, the patterns refer back to attributes that can explain why the items are misclassified. Inconsistent items can be found from the analysis of patterns.

4.3.1 Working mechanism of RCA procedure

In this section, we will briefly demonstrate how the RCA is applied to identify misclassified items. More details will be explained further in the Chapter 5 of numerical examples.

We use part of the first round result of LAD classification on AHP method dataset to illustrate the RCA procedure. Table 4.1 shows a sample of misclassified observations in testing. The unit price is US dollars and lead time is measured by days. 'Number of test' refers to how many times of an observation is tested during the 20 or 10 tests (one round of analysis).

Table 4.1: Misclassified observations of test (sample)

Test No.	Obs. No.	Part No.	Usage Rate	Unit Price	Lead Time	Orig. Class	Classes of Patterns Found	Class Classified by LAD	No. of Tests
5	13	601R75 100-209	0.3	1601.58	281	B	A, B	A	5
10							A, B	A	
13							A, B	A	
17							A, B	A	
18							A, B	A	
9	19	601R31 7094-1	0.05	113.74	260	B	zero	unclassified	3
15							A	A	

The two misclassified items are originally from Class B. We start with No. 13 observation. The patterns, which are created during the training stage of No. 5 test, are shown in Table 4.2. The attributes of No. 13 observation are usage rate 0.3, unit price US$1,602 and lead time 281 days, which conform to both Pattern 1 of Class A and Pattern 2 of Class B. Next we examine the weight of each pattern and use the weight to decide the class of observation. Since the Pattern 1 of Class A has a larger weight (1) than Pattern 2 of Class B's weight (0.4444), the observation is classified as Class A by LAD. Afterward, the No. 13 observation is tested 5 times over 20 tests and it is classified as Class A in all 5 tests by LAD, even though the original class is Class B. So the class of No. 13 observation changes from B to A.

No. 19 observation has the two problems of being misclassified as Class A and 'Zero'. Here 'Zero' means that LAD cannot find any patterns matched for this observation or there is an equal weight

of matched patterns of different classes, which is considered as unclassified. The data from the item does not provide enough information for LAD to determine its class. The No. 19 observation has been tested 3 times over 20 tests (Test No. 7, 9 and 15, respectively) which has the outcome of 2 misclassifications and 1 correct classification. The patterns found in the No. 7 Test are shown in Table 4.3. The attributes of the No. 19 observation are usage rate 0.05, unit price US$113.74 and lead time 260 days, which match the pattern 2 of Class B. The No. 19 observation is originally labelled Class B. So it is classified correctly in the No. 7 Test.

Table 4.2: Patterns created in the No. 5 Test

Class	Pattern		Weight
A	1	Unit Price Greater Than 270.265	1
		Lead Time Greater Than 207.5	
B	1	Unit Price Less Than 270.265	0.5556
		Lead Time Greater Than 139.5	
	2	Usage Rate Greater Than 0.065	0.4444
		Lead Time Greater Than 139.5	
C	1	Lead Time Less Than 139.5	1

The patterns found in the No. 9 Test are shown in Table 4.3. The attributes of No. 19 observation cannot match any patterns from Table 4.4. So the observation is considered as unclassified. The patterns found in the No. 15 Test are shown in Table 4.5. We can see that the No. 19 observation matches the pattern 1 of Class A. So it is misclassified in the No. 15 Test. Three tests receive three different results and show no consistent trend in the LAD tests, therefore, we keep this observation class unchanged.

Table 4.3: Patterns created in the No. 7 Test

Class	Pattern		Weight
A	1	Usage Rate Less Than 0.065	1
		Unit Price Greater Than 259.92	
		Lead Time Greater Than 207.5	
B	1	Usage Rate Greater Than 0.065	0.625
		Lead Time Greater Than 125.5	
	2	Lead Time Greater Than 239	0.375
C	1	Lead Time Less Than 125.5	1

Table 4.4: Patterns created in the No. 9 Test

Class	Pattern		Weight
A	1	Usage Rate Less Than 0.065	1
		Unit Price Greater Than 259.92	
		Lead Time Greater Than 207.5	
B	1	Usage Rate Greater Than 0.065	1
		Lead Time Greater Than 125.5	
C	1	Lead Time Less Than 125.5	1

Table 4.5: Patterns created in the No. 15 Test

Class	Pattern		Weight
A	1	Usage Rate Less Than 0.065	1
		Lead Time Greater Than 207.5	
B	1	Usage Rate Greater Than 0.065	1
		Lead Time Greater Than 125.5	
C	1	Lead Time Less Than 125.5	1

By following the procedure of RCA, the explanations are found for misclassification.

1. More than one pattern matched, but the pattern weight in another class is bigger than the patterns of the original class. In other words, the item shares more common attributes with the other class than with the original class.
2. Patterns are found in only other classes. There is inconsistency in the original class.
3. No pattern is matched in the existing class. The item does not give enough information for any of the classes.

The interpretable patterns make the explanation of misclassifications straightforward and easy to understand. In order to make sure that the change of class is not arbitrary and the inconsistency is corrected, the observation class is changed when it satisfies the following situation after an RCA procedure:

1. It is always misclassified in the one different class;
2. If there is no situation 1 found and test accuracy is less than 90%, we will need a special investigation for the misclassification. (See 5.1.4).

For instance, we see the No. 13 observation of Class B from Table 4.1, which is classified as Class A in all 5 tests, so the class of observation is changed from B to A. The RCA result is shown in Table 4.6.

Table 4.6: RCA result of misclassified observations

Obs. No.	Usage Rate	Unit Price	Lead Time	Orig. Class	Classes of Pattern Found	Class Classified by LAD	No. of Tests	Corrected Class
13	0.3	1601.58	281	B	A, B	A	5	A
19	0.05	113.74	260	B	zero	unclassified	3	no change
					A	A		

After inconsistent observations are corrected, the dataset is run through another round of training and tests. Misclassified observations are checked by the RCA procedure again to detect inconsistencies. The process of LAD classification and RCA is repeated until no more inconsistent observations are found. The RCA procedure is shown in Figure 4.1.

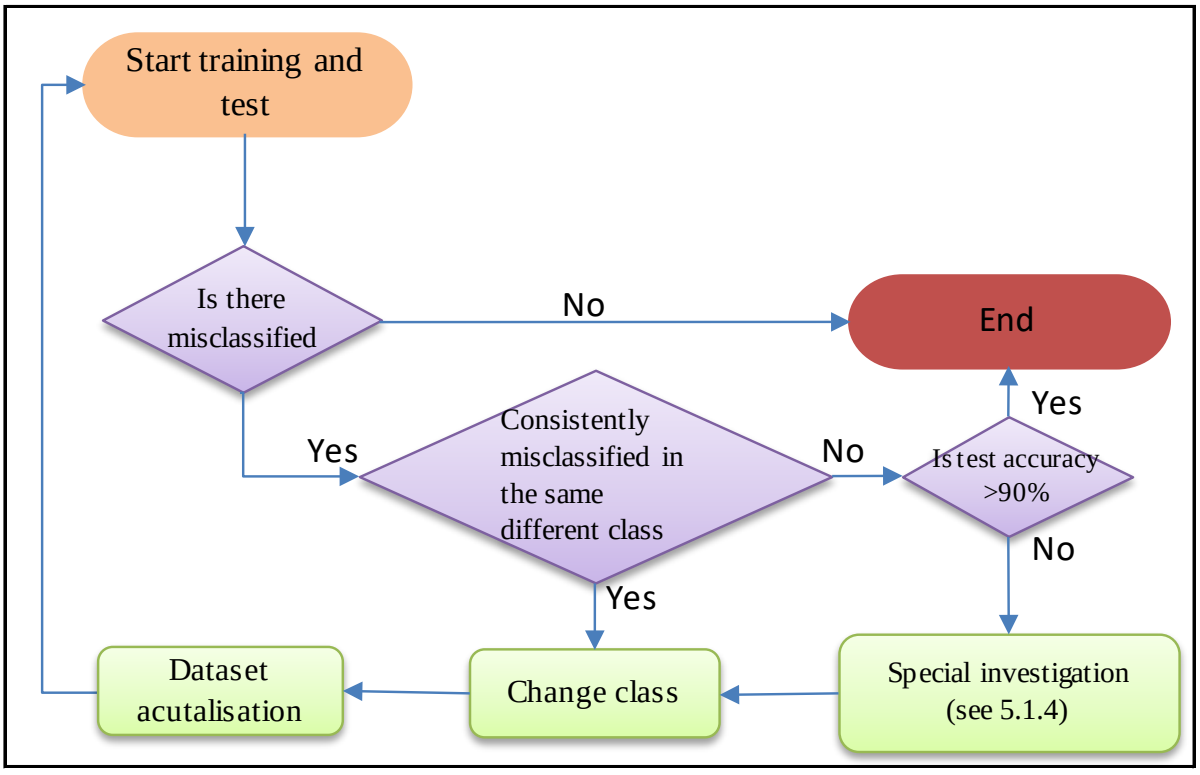

Figure 4.1: Root cause analysis procedure for inconsistency detection

CHAPTER 5 LAD CLASSIFICATION: NUMERICAL EXAMPLES

In this chapter, we use two sets of numerical examples to study the LAD classification applicability, capability and its effectiveness of detecting inconsistencies by combining it with the Root Cause Analysis (RCA) procedure. First, we examine the feasibility of LAD classification on spare parts inventory and then analyze the erroneous items of classification by applying the RCA procedure. Based on the results of RCA analysis, corresponding corrections will be made. Next, both the LAD classification and the RCA procedure will be run reiteratively until all inconsistency is corrected and an acceptable test accuracy (above 90%) is received. The second numerical example study is based on medical equipment inventory. Each numerical example includes one dataset of different ABC classification methods, namely AHP, DEA, DEA-like weighted linear optimization and scaled DEA-like weighted linear optimization.

5.1 Numerical example of spare parts inventory

5.1.1 Introduction

The dataset of LAD classification on inventory is adopted from the article by Rad, Shanmugarajan, and Wahab (2011). The dataset is a set of spare parts inventory from airlines. We are able to access part of the data containing 20 observations with three results of classification methods (see Table 5.1), namely Analytic Hierarchy Process (AHP), Data Envelopment Analysis (DEA) with Constant Return to Scale (CRS) and Variable Return to Scale (VRS).

As described in Chapter 4, we classify datasets with the LAD technique and then analyze the erroneous observations of classifications by applying the procedure of Root Cause Analysis (RCA). After making corrections, both the LAD classification and the RCA procedure will be run reiteratively until all inconsistencies are corrected and an acceptable test accuracy (above 90%) is received.

The dataset contains 20 observations with five attributes: Part Number, Usage Rate, Unit Price, Lead Time and Class. The attribute Class has three results for each of the three classification methods (AHP, VRS and CRS). The attribute Part Number has no effect on the classification and is only used for the purpose of identification. To make things easier, we use the observation number as the identification of an individual inventory item.

Table 5.1: Dataset organized by three ABC classification methods

Observation	Part No.	Usage Rate	Unit Price	Lead Time	ABC Classification		
					CRS	VRS	AHP
1	350689-7	0.05	4.16	98	A	A	C
2	J221P014	0.36	8.64	30	A	A	C
3	AS3582-038	0.21	2.42	197	A	A	B
4	AS3582-232	0.36	9.28	29	A	A	C
5	9452K71	0.05	0.8	50	A	A	C
6	M39029/58-363	0.03	4.48	98	A	A	C
7	AS3209-014	0.13	10.72	176	B	A	B
8	M39029/22-191	0.05	5.6	98	B	B	C
9	350690-7	0.08	7.04	24	B	A	C
10	NSA551607ND	0.05	28.16	98	B	B	C
11	CC670-38730-3	0.49	706.47	188	C	A	B
12	3E3291-1	0.15	1152.48	148	C	C	B
13	601R75100-209	0.3	1601.58	281	C	A	B
14	AS3582-228	0.26	9.12	50	C	A	C
15	SL618-3CM	0.03	6.9	260	C	A	B
16	601R31719-5	0.05	426.79	218	C	C	A
17	BA670-45691-25	0.03	19.79	39	C	C	C
18	49001-243	0.26	93.05	103	C	C	C
19	601R317094-1	0.05	113.74	260	C	B	B
20	601R40508-35	0.08	49.75	260	C	A	B

For the AHP classification method, the total of 20 observations consist of 1 observation of Class A, 8 observations of Class B and 11 observations of Class C. On VRS classification method, the dataset consists of 13 observations of Class A, 3 observations of Class B and 4 observations of Class C. On CRS classification method, the dataset consists of 6 observations of Class A, 4 observations of Class B and 10 observations of Class C.

The DEA methodology was presented initially by Charnes, Cooper and Rhodes (1978). The key ingredient is efficiency, which is defined as a ratio of weighted sum of outputs to a weighted sum of inputs, where the weight structure is calculated by means of mathematical programming, and constant returns to scale (CRS) are assumed. In 1984, Banker, Charnes and Cooper developed a model with variable returns to scale (VRS). Descriptions on AHP and DEA can be found in the Chapter 2 literature review.

5.1.2 LAD Classification on dataset

We use 80% of the dataset as a training subset and 20% as a testing subset. The partitions of training and testing subsets are random. One general rule is that we keep the proportions of the three classes in those subsets the same as in the original dataset, so we can achieve a more efficient analysis of the dataset.

The process of conducting LAD classification analysis follows the instructions described in Section 4.2 'LAD classification analysis procedure' in the previous chapter. After running twenty tests with randomly selecting training datasets and testing datasets, we calculate the average test accuracy. We consider these twenty tests as one round analysis of LAD classification. The confusion matrix presents the overall accuracy of the first round LAD analysis (shown in Table 5.2).

Table 5.2: The average test accuracy of the first round LAD analysis

Classification	Class/Predicted	A	B	C	Unclassified	Total	Average Test Accuracy
AHP	A	NA	NA	NA	NA	NA	(30+40)/(40+40) = 87.5%
	B	8	30	0	2	40	
	C	0	0	40	0	40	
VRS	A	53	3	1	3	60	(53+10+8)/(60+20+20) =71%
	B	3	10	6	1	20	
	C	5	5	8	2	20	
CRS	A	16	0	2	2	20	(16+5+25)/(20+20+40) =57.5%
	B	5	5	8	2	20	
	C	5	7	25	3	40	

The numbers in Table 5.2 stand for how many observations are being tested during one round of analysis. If an observation is tested twice, it will count as 2, and so on. The 'NA' (Not Applicable) in Table 5.2 refers to the AHP dataset that has only one observation in Class A, so we are not able to test the Class A pattern at this time. This may be part of the reason that the LAD classification on AHP dataset reaches the best average test accuracy of 87.50%, followed by 71% for VRS dataset and 57.50% for CRS dataset. Next, we conduct the RCA procedure to detect any inconsistencies and to achieve better test accuracy (as described in section, 5.1.3.).

5.1.3 Misclassification analysis

The first round of analysis of the LAD classification has averages of 12.5%, 29% and 42.5% of misclassification on AHP, VRS and CRS datasets, respectively. The misclassified observations on AHP, VRS and CRS datasets are listed in Table 5.3, Table 5.4 and Table 5.5, respectively. The column 'Ratio of Misclassification' shows the detail of misclassification. '5/5' means 5 misclassifications out of a total of 5 test runs in this round of analysis, and so on. The same is true for other tables below.

Table 5.3: Misclassified observations from 1st round analysis of the AHP dataset

Observation from table 5.1	Original Class	Classes of Pattern Found	Class Classified by LAD	Ratio of Misclassification
13	B	A, B	A	5/5
15	B	zero	unclassified	1/4
		A	A	1/4
19	B	zero	unclassified	1/3
		A	A	1/3
4	B	A, B	A	1/7

Table 5.4: Misclassified observations from 1st round analysis of the VRS dataset

Observation from table 5.1	Original Class	Classes of Pattern Found	Class Classified by LAD	Ratio of Misclassification
3	A	A, C	C	1/8
8	B	A	A	3/3
9	A	A, B	B	1/5
		zero	unclassified	1/5
12	C	A	A	1/8
13	A	zero	unclassified	1/3
15	A	B	B	1/2
		zero	unclassified	1/2
16	C	B	B	3/4
17	C	B	B	2/5
		zero	unclassified	2/5
		A	A	1/5
18	C	A	A	3/3
19	B	zero	unclassified	1/7
		C	C	4/7
		A, C	C	2/7
20	A	B	B	1/3

Table 5.5: Misclassified observations from 1st round analysis of the CRS dataset

Observation from table 5.1	Original Class	Classes of Pattern Found	Class Classified by LAD	Ratio of Misclassification
3	A	zero	unclassified	1/3
4	A	C	C	2/3
		zero	unclassified	1/3
7	B	C	C	5/7
		zero	unclassified	2/7
8	B	A	A	4/4
9	B	A	A	1/2
10	B	C	C	3/7
12	C	B	B	2/5
14	C	A	A	5/5
15	C	zero	unclassified	3/3
17	C	B	B	5/5

Next, we conduct the procedure of RCA for misclassified observations by checking for any contradictions or repetitions of observations.

In the AHP classification dataset, there is no contradiction or repetition found in the training subsets or testing subsets. Next, we examine for any possibility of inconsistency items in the dataset. Three observations are misclassified or unclassified in the first run analysis. The column of 'Ratio of Misclassification' in the above Tables shows the detail of misclassification. For example, the No. 13 observation in Table 5.3 is marked as '5/5', which means 5 misclassifications out of a total of 5 test runs in the first run analysis, and so on. We also know that the No. 13 observation is misclassified as Class A by LAD, which originally belonged to Class B in all five tests.

Table 5.6: RCA result of 1st round analysis of the AHP dataset

Observation from table 5.1	Original Class	Classes of Pattern Found	Class Classified by LAD	Ratio of Misclassification	Corrected Class
13	B	A, B	A	5/5	A
15	B	zero	unclassified	1/4	no change
		A	A	1/4	
19	B	zero	unclassified	1/3	no change
		A	A	1/3	

The No. 15 observation is misclassified as Class A once (1/4), unclassified once (1/4) by LAD, which originally belonged to Class B and classified twice (2/4) correctly in four tests. The No. 19 observation is misclassified as Class A once (1/3), unclassified once (1/3) by LAD that originally belonged to Class B and classified once (1/3) correctly in three tests. Based on the rule of RCA,

the class of No. 13 observation is changed from Class B to A, while the class of No. 15 and 19 observations are kept unchanged. The RCA result is shown in Table 5.6.

The same process is applied to an analysis of the first round of misclassifications by LAD on a VRS and CRS dataset. The RCA results are presented in Table 5.7 and Table 5.8, respectively.

Table 5.7: RCA results of 1st round analysis of the VRS dataset

Observation from table 5.1	Original Class	Classes of Pattern Found	Class Classified by LAD	Ratio of Misclassification	Corrected Class
3	A	A, C	C	1/8	no change
8	B	A	A	3/3	A
9	A	A, B	B	1/5	no change
		zero	unclassified	1/5	
12	C	A	A	1/8	no change
13	A	zero	unclassified	1/3	no change
15	A	B	B	1/2	no change
		zero	unclassified	1/2	
16	C	B	B	3/4	no change
17	C	B	B	2/5	no change
		zero	unclassified	2/5	
		A	A	1/5	
18	C	A	A	3/3	A
19	B	zero	unclassified	1/7	no change
		C	C	4/7	
		A, C	C	2/7	
20	A	B	B	1/3	no change

Table 5.8: RCA result of 1st round analysis of the CRS dataset

Observation from table 5.1	Original Class	Classes of Pattern Found	Class Classified by LAD	Ratio of Misclassification	Corrected Class
3	A	zero	unclassified	1/3	no change
4	A	C	C	2/3	no change
		zero	unclassified	1/3	
7	B	C	C	5/7	no change
		zero	unclassified	2/7	
8	B	A	A	4/4	A
9	B	A	A	1/2	no change
10	B	C	C	3/7	no change
12	C	B	B	2/5	no change
14	C	A	A	5/5	A
15	C	zero	unclassified	3/3	no change
17	C	B	B	5/5	B

After conducting the procedure of RCA and making corrections on inconsistent observations, we run the LAD classification again on the ‘new’ dataset, which is considered as the start of the second

round of data analysis. The process of data analysis is repeated until no more inconsistency items are found. On the AHP classification dataset, the process is repeated until a fifth round of analysis.

5.1.4 Special investigation for misclassification

During the third-round analysis, only unclassified observations are found in the testing, which means the RCA procedure cannot provide more information on inconsistencies. The test accuracy reaches 88.75% after a third round of analysis that is still below the normal expectations of more than 90%. Therefore, we intend to explore this further to find other possibilities.

The No. 3 observation is unclassified twice and classified correctly four times over six tests during the third-round analysis, while the No. 11 observation is unclassified seven times over seven tests.

Firstly, we use the leave-one-out cross validation to test the observation again. The results show that the No. 3 observation is classified correctly and the No. 11 observation is unclassified. We will not change the class of the No. 3 observation. So further analysis is needed on the No. 11 observation.

Next we use a MATLAB 3D visualization to help us make a decision. As shown in Figure 5.1, the No. 3 observation is closest to Class B, which conforms to our judgement. The No. 11 observation is not very far from Class A and B observations, but it cannot be Class B based on the results of the LAD analysis. It cannot be Class C either from the data visualisation. In the real industrial world, inventory items should be in certain classes. So we change the No. 11 observation from Class B to A. Even if the change is wrong, the LAD technique and RCA procedure can still correct it in the latter round of analysis.

To diminish the human factor as much as possible, a special investigation is initiated only when the situation satisfies all the following criteria at the same time.

1) Unclassified observations are found in the tests

2) The tests of observation are all unclassified;

3) The average test accuracy is below expectations (normally above 90%).

Please note that in our experiments, the special investigation is implemented only once. This special investigation is an isolated incident and unnecessary for the misclassification analysis.

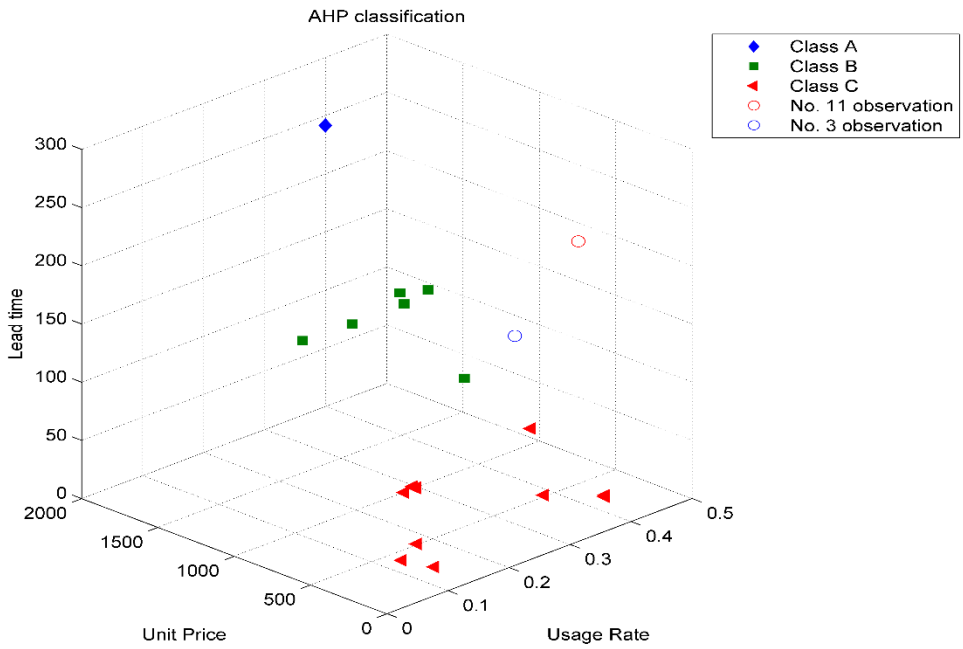

Figure 5.1: Visualisation of observations on the AHP dataset

5.1.5 Analysis results

We conduct six rounds of LAD classification and RCA procedures for correcting the inconsistent observations. The final average test accuracy on AHP dataset reaches 95.00%. The results of data analysis by LAD classification on AHP dataset are shown in the Table 5.9.

The same process is applied to VRS dataset to identify inconsistent observations and to make corrections by combining with the RCA procedure. The data analysis has been implemented in seven rounds for VRS dataset. The corresponding results of LAD classification on VRS dataset are displayed in Table 5.10 and 5.11.

After making corrections on inconsistency in datasets, average test accuracy has been improved from 87.50% to 91.25% in five rounds analysis for the AHP dataset, from 71% to 100% in seven rounds analysis for the VRS dataset.

Table 5.9: Results of LAD analysis with RCA of the AHP dataset

Round	Obs. from table 5.1	Orig. Class	Pattern Found	Class Classified by LAD	Ratio of Misclassification	Corrected Class	Average Accuracy
2nd	16	A	B	B	11/11	B	81.25%
	11	B	Zero	Unclassified	3/3	no change	
	19	B	Zero	Unclassified	1/2	no change	
			B	B	1/2		
3rd	3	B	Zero	Unclassified	2/6	no change	88.75%
	11	B	Zero	Unclassified	7/7	A	
4th	11	A	Zero	Unclassified	9/9	no change	75.00%
	13	A	Zero	Unclassified	7/11	no change	
	12	B	A, B	A	4/4	A	
5th	11	A	Zero	Unclassified	3/3	no change	91.25%
	3	B	Zero	Unclassified	3/3	no change	
	7	B	Zero	Unclassified	3/3	no change	
	16	B	A, B	A	2/3	no change	

Table 5.10: Results of LAD analysis with RCA of the VRS dataset

Round	Obs. from table 5.1	Orig. Class	Class of Pattern Found	Class Classified by LAD	Ratio of Misclassification	Corrected Class	Average Accuracy
2nd	13	A	zero	unclassified	1/4	no change	54.00%
			B	B	2/4		
	18	A	A, B	B	3/3	B	
	10	B	A, C	C	9/9	C	
	19	B	zero	unclassified	7/11	no change	
			C	C	2/11		
			A	A	1/11		
	12	C	A	A	8/8	A	
	16	C	B	B	5/7	no change	
			A	A	2/7		
	17	C	B	B	5/5	B	
3rd	10	C	A, B	B	3/10	no change	60.00%
			A	A	7/10		
	11	A	zero	unclassified	2/5	no change	
	13	A	C	C	4/4	C	
	16	C	zero	unclassified	3/10	no change	
			B	B	1/10		
	17	B	C	C	1/8	no change	
			A	A	7/8		
	18	B	A, C	C	3/6	no change	
			A	A	3/6		
	19	B	C	C	6/6	C	

Table 5.11:Results of LAD analysis with RCA of the VRS dataset

Round	Obs. from table 5.1	Orig. Class	Class of Pattern Found	Class Classified by LAD	Ratio of Misclassification	Corrected Class	Average Accuracy
4th	10	C	zero	unclassified	2/5	no change	63.00%
			A	A	3/5		
	11	A	zero	unclassified	2/6	no change	
	12	A	C	C	3/3	C	
	13	C	zero	unclassified	1/4	no change	
			A	A	3/4		
	17	B	C	C	8/10	no change	
			A	A	2/10		
	18	B	A	A	10/10	A	
	20	A	zero	unclassified	1/5	no change	
			C	C	2/5		
5th	6	A	A, B	B	3/3	B	72.50%
	10	C	A	A	4/4	A	
	11	A	C	C	3/3	C	
	12	C	A	A	1/5	no change	
	13	C	zero	unclassified	4/4	no change	
	15	A	A, B	B	3/3	B	
	18	A	A, C	C	1/4	no change	
	20	A	A, C	C	2/3	no change	
			zero	unclassified	1/3		
6th	8	A	zero	unclassified	1/4	no change	90.00%
	18	A	C	C	3/4	no change	
	20	A	zero	unclassified	1/3	no change	
	19	C	A	A	3/3	A	
7th	no misclassified						100%

For the CRS dataset, we follow the same process to identify inconsistent observations and to make corrections by combining with the RCA procedure. The data analysis has been carried out four rounds for the CRS dataset. The corresponding results of LAD classification on CRS dataset is displayed in Table 5.12.

After making corrections on inconsistency in CRS dataset, average test accuracy has been improved from 57.50%，72%，86% to 92.00% in four rounds analysis for the CRS dataset.

In summary, inconsistencies in three datasets have been corrected by the LAD technique along with the RCA procedure. The test accuracy of each dataset has been improved greatly, specifically from 87.5% to 91.25%, from 71% to 100%, and from 57.5% to 92% for the AHP, VRS and CRS datasets, respectively.

Table 5.12: Results of LAD analysis with RCA of the CRS dataset

Round	Obs. from table 5.1	Orig. Class	Class of Pattern Found	Class Classified by LAD	Ratio of Misclassification	Corrected Class	Average Accuracy
2nd	2	A	zero	unclassified	1/7	no change	72.00%
	3	A	zero	unclassified	1/6	no change	
			C	C	3/6	no change	
	4	A	zero	unclassified	1/4	no change	
	7	B	A, C	A	4/5	no change	
			C	C	1/5		
	9	B	A	A	5/5	A	
	12	C	B	B	4/6	no change	
	15	C	A	A	3/3	A	
	18	C	zero	unclassified	2/4	no change	
			B	B	1/4		
	20	C	B	B	2/3	no change	
3rd	7	B	A	A	6/6	A	86.00%
	12	C	B	B	2/6	no change	
	18	C	B, C	B	2/6	no change	
	20	C	B	B	1/4	no change	
			zero	unclassified	3/4		
4th	10	B	zero	unclassified	5/5	no change	92.00%
	20	C	zero	unclassified	3/3	no change	

5.2 Numerical example of medical equipment inventory

5.2.1 Introduction

To further investigate the effectiveness of LAD classification and the RCA procedure for improving the consistency of classification and test accuracy, the second set numerical example is introduced. The dataset of the example is used by many researchers and considered one of most common used datasets to demonstrate classification techniques in an inventory classification problem (Hadi-Vencheh, 2010; Ng, 2007; Ramanathan, 2006; Yu, 2011). The dataset is a spare parts inventory of medical equipment with 47 SKUs. The datasets are classified by three classification techniques, namely Analytic Hierarchy Process (AHP), DEA-like weighted linear optimization and scaled DEA-like weighted linear optimization. Those classification techniques are proposed by Flores, Olson, & Dorai (1992), Ramanathan (2006) and Ng (2007) respectively.

First, we apply the LAD machine learning technique to classify the dataset, to list misclassified observations and to compute the test accuracy. Then, the RCA procedure is implemented to detect

inconsistencies. According to the results of the RCA analysis, the corrections are made. After that, the LAD classification process is executed again as well as the RCA procedure until inconsistent observations are fixed and the test accuracy is acceptable.

5.2.2 AHP dataset

The Analytic Hierarchy Process (AHP) classification of this dataset is proposed by Flores, Olson, & Dorai (1992). They use average unit cost, annual dollar usage, criticality and lead time as attributes to analyse contributions of the observations. Criticality has three values, which are 1 for a very critical item, 0.50 for a moderately critical item and 0.01 for a relatively less critical item. Lead time varies from 1 to 7 weeks.

In the next stage, we implement the LAD classification on this AHP dataset by randomly partitioning 80% and 20% of the dataset as the training subset and testing subset. After 10 tests, the average test accuracy is 69%. The misclassified observations are shown in Table 5.13.

Table 5.13: The misclassified items of 1st round analysis of the AHP dataset

SKU	Original Class	Classes of Pattern Found	Class Classified by LAD	Ratio of Misclassification
3	C	A	A	3/3
5	B	C	C	3/3
11	B	C	C	3/4
		A	A	1/4
12	B	C	C	1/2
14	B	A, B	A	1/4
16	C	B, C	B	2/2
17	B	C	C	1/3
		Zero	Unclassified	1/3
18	A	B	B	3/3
21	A	B, C	B	2/2
28	C	B	B	1/2
29	B	C	C	1/2
		Zero	Unclassified	1/2
38	C	B, C	B	2/2
40	C	Zero	Unclassified	1/2
45	B	C	C	2/2

The RCA procedure is implemented based on the test results and the following change is made. The No. 3 SKU is changed from Class C to A, the No. 5 and 45 SKUs from Class B to C, the No. 16 and 38 SKUs from Class C to Class B, the No. 18 and 21 SKUs from Class A to Class B. The analysis decision of the RAC procedure is shown in Table 5.14.

Table 5.14: RCA results of 1st round analysis of the AHP dataset

SKU	Original Class	Classes of Pattern Found	Class Classified by LAD	Ratio of Misclassification	Corrected Class
3	C	A	A	3/3	A
5	B	C	C	3/3	C
11	B	C	C	3/4	No change
		A	A	1/4	
12	B	C	C	1/2	No change
14	B	A, B	A	1/4	No change
16	C	B, C	B	2/2	B
17	B	C	C	1/3	No change
		Zero	Unclassified	1/3	
18	A	B	B	3/3	B
21	A	B, C	B	2/2	B
28	C	B	B	1/2	No change
29	B	C	C	1/2	No change
		Zero	Unclassified	1/2	
38	C	B, C	B	2/2	B
40	C	Zero	Unclassified	1/2	No change
45	B	C	C	2/2	C

Next, we start the next rounds of LAD classification and RCA procedure on the ‘new’ dataset until no inconsistency is found. The process is implemented in four rounds and the final test accuracy reaches 94%. The results of each round of analysis are shown in Table 5.15 and the change in test accuracy is presented in Figure 5.2.

Table 5.15: Results of LAD Classification with RCA of the AHP dataset

Round	SKU	Orig. Class	Class of Pattern Found	Class Classified by LAD	Ratio of Misclassification	Corrected Class	Test Accuracy
2nd	3	A	Zero	Unclassified	3/3	No change	77%
	11	B	A	A	2/3	No change	
	14	B	Zero	Unclassified	1/2	No change	
	10	B	Zero	Unclassified	1/1	No change	
	15	A	A, B	B	1/2	No change	
	21	B	A, B	A	3/3	A	
	24	A	B	B	2/3	No change	
	26	C	Zero	Unclassified	2/4	No change	
	29	B	C	C	3/3	C	
	31	B	Zero	Unclassified	3/3	No change	
	32	B	A, B	A	1/2	No change	
	36	B	A, B	A	1/2	No change	
3rd	1	A	Zero	Unclassified	1/1	No change	85%
	4	C	Zero	Unclassified	3/3	No change	
	11	B	A	A	2/2	A	
	12	B	B, C	C	1/2	No change	
	17	B	B, C	C	1/2	No change	
	18	B	Zero	Unclassified	1/2	No change	
	31	B	Zero	Unclassified	1/3	No change	
	32	B	A, B	A	2/4	No change	
			Zero	Unclassified	1/4		
	36	B	A, B	A	2/4	No change	
4th	24	A	Zero	Unclassified	1/1	No change	94%
	10	B	Zero	Unclassified	2/2	No change	
	32	B	Zero	Unclassified	1/2	No change	
			A	A	1/2		
	36	B	A	A	1/2	No change	

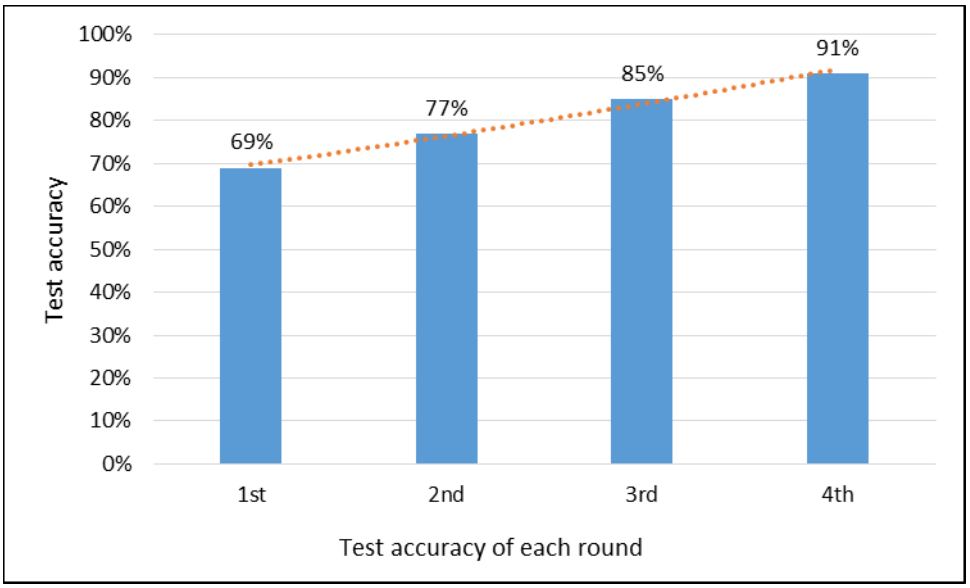

Figure 5.2: The change of test accuracy on the AHP dataset

5.2.3 Optimal dataset

Ramanathan (2006) proposed a multi-criteria weighted linear optimization model to classify inventory. They formulated a DEA-like linear optimization function to determine optimal scores for all inventory items. We call the result of classification the 'Optimal' dataset. The detail of the proposed model can be found in the article (Ramanathan, 2006).

In the next step, we conduct the LAD classification on this Optimal dataset by randomly partitioning 80% and 20% of the dataset as the training subset and testing subset. After 10 tests, the average test accuracy is 47%. The misclassifications are shown in Table 5.16.

The RCA procedure is implemented based on the LAD test results and the following change is made. The No. 6 SKU is changed from Class A to C, the No. 33 SKU from Class A to B, the No. 4, 5, 14 and 18 SKUs from Class A to Class B, the No. 23, 30, 36, 39 and 41 SKUs from Class C to Class B, the No. 7 and 10 SKUs from Class C to A, the No. 27 and 40 SKUs from Class C to Class B. The change decision of the RAC procedure is shown in Table 5.17.

Table 5.16: Misclassified observations from 1st round analysis of the Optimal dataset

SKU	Original Class	Classes of Pattern Found	Class Classified by LAD	Ratio of Misclassification
2	A	Zero	Unclassified	3/3
4	B	A	A	2/2
5	B	A	A	1/1
6	A	C	C	2/2
7	C	A	A	2/2
8	B	C	C	1/2
9	B	A	A	2/4
10	C	A	A	1/1
13	A	B	B	1/2
14	B	A	A	2/2
15	C	B	B	1/2
18	B	A	A	3/3
20	C	B	B	1/4
22	C	B	B	1/2
23	B	C	C	2/2
25	C	B	B	1/3
27	C	B	B	2/2
29	A	B	B	1/3
		Zero	Unclassified	1/3
30	B	C	C	3/3
32	C	B	B	1/2
33	A	B	B	2/2
34	A	Zero	Unclassified	1/1
35	C	Zero	Unclassified	1/2
36	B	C	C	2/2
38	C	Zero	Unclassified	1/2
		B	B	1/2
39	B	C	C	1/1
40	C	B	B	1/1
41	B	C	C	1/1
45	B	A	A	2/3
		C	C	1/3
47	C	Zero	Unclassified	2/3

Next we start a second round of LAD classification and RCA procedure on the ‘new’ dataset. The misclassified observations and RCA results are shown in Table 5.18. The test accuracy has improved slightly to 58%.

Table 5.17: RCA results of 1st round analysis of the Optimal dataset

SKU	Original Class	Class of Pattern Found	Class Classified by LAD	Ratio of Misclassification	Corrected Class
4	B	A	A	2/2	A
5	B	A	A	2/2	A
6	A	C	C	1/1	C
7	C	A	A	3/3	A
10	C	A	A	1/1	A
14	B	A	A	1/1	A
18	B	A	A	2/2	A
23	B	C	C	1/1	C
27	C	B	B	2/2	B
30	B	C	C	3/3	C
33	A	B	B	2/2	B
36	B	C	C	3/3	C
39	B	C	C	2/2	C
40	C	B	B	1/1	B
41	B	C	C	1/1	C

Table 5.18: Misclassification and RCA results of 2nd round analysis on Optimal dataset

SKU	Original Class	Class of Pattern Found	Class Classified by LAD	Ratio of Misclassification	Corrected Class
1	A	Zero	Unclassified	1/2	No change
4	A	Zero	Unclassified	3/3	No change
5	A	C	C	2/2	C
6	C	A	A	2/2	A
7	A	C	C	3/3	C
8	B	A, C	A	1/1	A
9	B	A, B	A	1/2	No change
		Zero	Unclassified	1/2	
10	A	Zero	Unclassified	2/3	No change
11	C	A	A	1/2	No change
12	A	Zero	Unclassified	1/2	No change
13	A	B	B	2/2	B
15	C	B	B	2/2	B
17	C	A, C	A	2/2	A
18	A	B	B	2/4	No change
23	C	B	B	1/3	No change
		Zero	Unclassified	2/3	
27	B	C	C	1/1	C
28	B	A	A	1/5	No change
29	A	B	B	2/2	B
33	B	C	C	1/3	No change
34	A	B	B	1/1	B
39	C	B	B	3/3	B
40	B	C	C	1/4	No change
		Zero	Unclassified	1/4	
45	B	A	A	2/3	No change
		Zero	Unclassified	1/3	

The process is implemented in eight rounds on the Optimal dataset until no inconsistency is found. The final test accuracy reaches 91%. The detail of the third to eighth round analysis results are presented in Table 5.19-5.24 and Figure 5.3 shows the change of test accuracy during the inconsistency correction process.

Table 5.19: Misclassification and RCA results of 3rd round analysis on Optimal dataset

SKU	Original Class	Class of Pattern Found	Class Classified by LAD	Ratio of Misclassification	Corrected Class	Test Accuracy
2	A	Zero	Unclassified	1/3	No change	73.33%
4	A	Zero	Unclassified	1/2	No change	
5	C	A	A	1/1	A	
6	A	C	C	2/2	C	
7	C	A	A	2/3	No change	
9	B	A	A	2/2	A	
13	B	A	A	1/1	A	
14	A	A, B	B	2/2	B	
15	B	C	C	1/1	C	
18	A	B	B	1/1	B	
23	C	Zero	Unclassified	1/4	No change	
28	B	Zero	Unclassified	1/3	No change	
33	B	C	C	1/1	C	
34	B	Zero	Unclassified	3/3	No change	
37	C	B, C	B	1/3	No change	
39	B	Zero	Unclassified	1/2	No change	
45	B	Zero	Unclassified	3/4	No change	

Table 5.20: Misclassification and RCA results of 4th round analysis on Optimal dataset

SKU	Original Class	Class of Pattern Found	Class Classified by LAD	Ratio of Misclassification	Corrected Class	Test Accuracy
4	A	Zero	Unclassified	1/2	No change	77.78%
5	A	C	C	2/2	C	
8	A	A, C	C	1/2	No change	
10	A	C	C	1/2	No change	
14	B	A	A	3/3	A	
17	A	C	C	1/2	No change	
18	B	A	A	3/3	A	
19	A	C	C	1/2	No change	
		B	B	1/2		
23	C	Zero	Unclassified	1/1	No change	
33	C	B, C	B	2/2	B	
34	B	Zero	Unclassified	1/2	No change	
37	C	B, C	B	1/3	No change	
43	C	B, C	B	1/3	No change	

Table 5.21: Misclassification and RCA results of 5th round analysis on Optimal dataset

SKU	Original Class	Class of Pattern Found	Class Classified by LAD	Ratio of Misclassification	Corrected Class	Test Accuracy
1	A	C	C	1/3	No change	80.00%
		Zero	Unclassified	1/3		
4	A	C	C	1/3	No change	
		Zero	Unclassified	2/3		
5	C	A, C	A	1/1	A	
15	C	A, C	A	1/3	No change	
28	B	Zero	Unclassified	4/4	No change	
29	B	Zero	Unclassified	1/3	No change	
31	B	Zero	Unclassified	1/1	No change	
37	C	B, C	B	1/2	No change	
45	B	Zero	Unclassified	3/3	No change	
47	C	Zero	Unclassified	2/2	No change	

Table 5.22: Misclassification and RCA results of 6th round analysis on Optimal dataset

SKU	Original Class	Class of Pattern Found	Class Classified by LAD	Ratio of Misclassification	Corrected Class	Test Accuracy
4	A	Zero	Unclassified	1/2	No change	83.00%
5	A	C	C	3/3	C	
23	C	Zero	Unclassified	1/2	No change	
28	B	Zero	Unclassified	2/3	No change	
34	B	Zero	Unclassified	2/2	No change	
37	C	B, C	B	3/3	B	
39	B	Zero	Unclassified	2/3	No change	
45	B	Zero	Unclassified	3/4	No change	

Table 5.23: Misclassification and RCA results of 7th round analysis on Optimal dataset

SKU	Original Class	Class of Pattern Found	Class Classified by LAD	Ratio of Misclassification	Corrected Class	Test Accuracy
1	A	C	C	1/2	No change	84.00%
		Zero	Unclassified	1/2		
4	A	Zero	Unclassified	2/2	No change	
28	B	Zero	Unclassified	3/3	No change	
34	B	Zero	Unclassified	2/4	No change	
43	C	B	B	3/3	B	
45	B	Zero	Unclassified	2/2	No change	
47	C	B	B	2/2	B	

Table 5.24: Misclassification and RCA results of 8th round analysis on Optimal dataset

SKU	Original Class	Class of Pattern Found	Class Classified by LAD	Ratio of Misclassification	Corrected Class	Test Accuracy
1	A	Zero	Unclassified	1/2	No change	91.00%
4	A	Zero	Unclassified	1/1	No change	
8	A	C	C	1/2	No change	
11	C	Zero	Unclassified	1/2	No change	
17	A	C	C	1/3	No change	
19	A	Zero	Unclassified	1/3	No change	
28	B	Zero	Unclassified	3/3	No change	

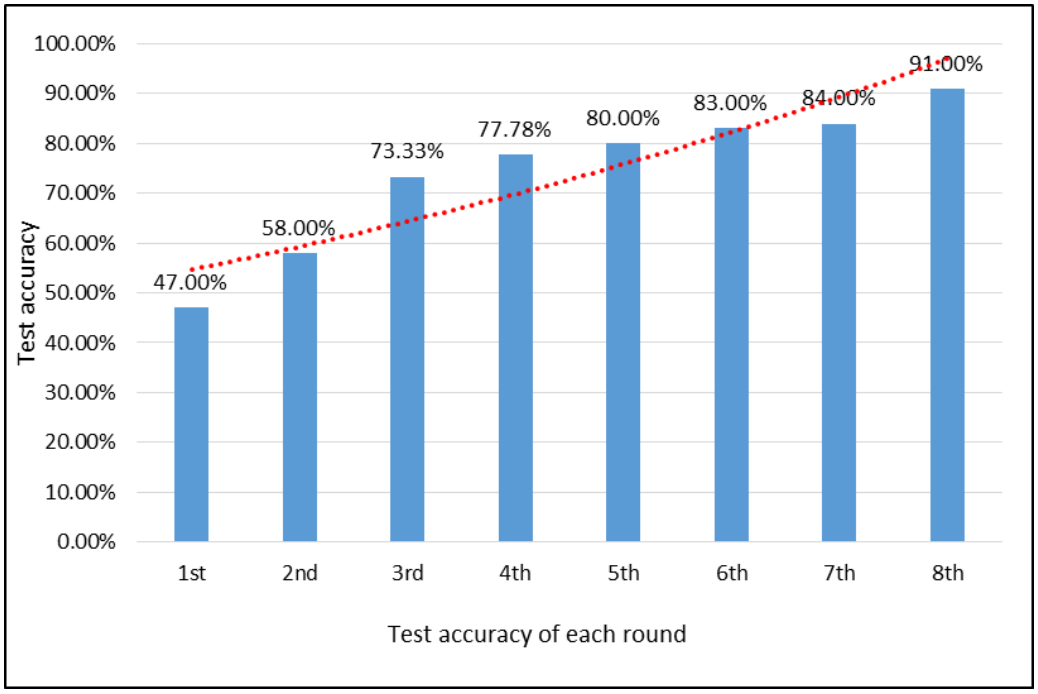

Figure 5.3: The change of test accuracy on the Optimal dataset

5.2.4 Scaled dataset

Ng (2007) presented a scaled DEA-like weighted linear optimization which is similar to the classification scheme proposed by Ramanathan (2006). The main difference is that Ng (2007) converted all measurements into a 0-1 scale before formulating the linear model. In addition, the attribute of a critical factor is not used due to its nominal data and being discontinuous. The dataset used is same as Flores, Olson, & Dorai (1992) and Ramanathan (2006). We call it a 'Scaled' dataset. The detail of this method can be found in the article (Ng, 2007).

Next step, we implement the LAD classification on this Scaled dataset by randomly partitioning the 80% and 20% of the dataset as the training subset and testing subset. After 10 tests, the average test accuracy is 55%. The misclassified observations are shown in Table 5.25. Following the procedure of RCA, we make the correction of each observations' class, and only those observations with changes are shown in Table 5.26.

Table 5.25: Misclassified observations of 1st round analysis on the Scaled dataset

SKU	Original Class	Classes of Pattern Found	Class Classified by LAD	Ratio of Misclassification
6	A	B	B	1/1
7	B	A	A	1/1
8	B	A	A	4/4
9	A	B	B	1/3
		Zero	Unclassified	1/3
11	C	B	B	1/1
12	B	C	C	2/2
14	B	A	A	1/4
		Zero	Unclassified	3/4
15	C	Zero	Unclassified	2/4
16	C	Zero	Unclassified	1/1
18	C	B	B	1/2
19	B	C	C	1/1
23	B	A	A	1/2
		Zero	Unclassified	1/2
27	C	B	B	1/1
29	A	B	B	1/1
31	B	C	C	2/3
		Zero	Unclassified	1/3
32	C	B	B	1/2
33	B	C	C	2/2
34	B	C	C	3/3
35	C	B	B	1/2
39	B	C	C	2/2
40	B	C	C	2/3
		Zero	Unclassified	1/3
44	C	Zero	Unclassified	2/2
45	B	C	C	3/3
46	C	Zero	Unclassified	1/1

Table 5.26: RCA results of the 1st round analysis on the Scaled dataset

SKU	Original Class	Class of Pattern Found	Class Classified by LAD	Ratio of Misclassification	Corrected Class
6	A	B	B	1/1	B
7	B	A	A	1/1	A
8	B	A	A	4/4	A
11	C	B	B	1/1	B
12	B	C	C	2/2	C
19	B	C	C	1/1	C
27	C	B	B	1/1	B
29	A	B	B	1/1	B
33	B	C	C	2/2	C
34	B	C	C	3/3	C
39	B	C	C	2/2	C
45	B	C	C	3/3	C

Next, we start a second round of LAD classification and RCA procedure on the 'new' Scaled dataset. The misclassified observations and RCA results are shown in Table 5.27. The accuracy of the test has improved to 72.22%.

Table 5.27: Misclassification and RCA result of 2nd round analysis on the Scaled dataset

SKU	Original Class	Class of Pattern Found	Class Classified by LAD	Ratio of Misclassification	Corrected Class
1	A	Zero	Unclassified	1/1	No change
3	A	Zero	Unclassified	2/2	No change
6	B	A	A	1/2	No change
		Zero	Unclassified	1/2	
7	A	Zero	Unclassified	1/3	No change
		B	B	2/3	
8	A	Zero	Unclassified	1/3	No change
11	B	A	A	1/3	No change
		C	C	2/3	
12	C	A	A	1/1	A
13	A	B	B	2/2	B
14	B	A	A	1/2	No change
		Zero	Unclassified	1/2	
15	C	B	B	2/2	B
23	B	Zero	Unclassified	2/4	No change
31	B	C	C	2/2	C
40	B	C	C	2/2	C

The LAD classification and RCA process is implemented for a total of four rounds on the Scaled dataset until no inconsistency is found. The final test accuracy reaches 93.33%. The results of the

third and fourth round of analysis are shown in Table 5.28 and 5.29. The change of test accuracy is presented in Figure 5.4.

Table 5.28: Misclassification and RCA result of 3rd round analysis on the Scaled dataset

SKU	Original Class	Class of Pattern Found	Class Classified by LAD	Ratio of Misclassification	Corrected Class	Test Accuracy
1	A	Zero	Unclassified	1/1	No change	83.33%
6	B	A	A	3/3	A	
7	A	A, B	B	1/2	No change	
9	A	A, B	B	1/2	No change	
11	B	A	A	1/1	A	
12	A	B	B	1/1	B	
15	B	Zero	Unclassified	1/1	No change	
16	C	B	B	1/1	B	
27	B	Zero	Unclassified	1/3	No change	
29	B	Zero	Unclassified	2/3	No change	
31	C	B	B	2/2	B	

Table 5.29: Misclassification and RCA result of 4th round analysis on the Scaled dataset

SKU	Original Class	Class of Pattern Found	Class Classified by LAD	Ratio of Misclassification	Corrected Class	Test Accuracy
12	B	Zero	Unclassified	2/2	No change	93.33%
16	B	Zero	Unclassified	2/2	No change	
31	B	Zero	Unclassified	1/3	No change	
17	C	B	B	1/1	No change	

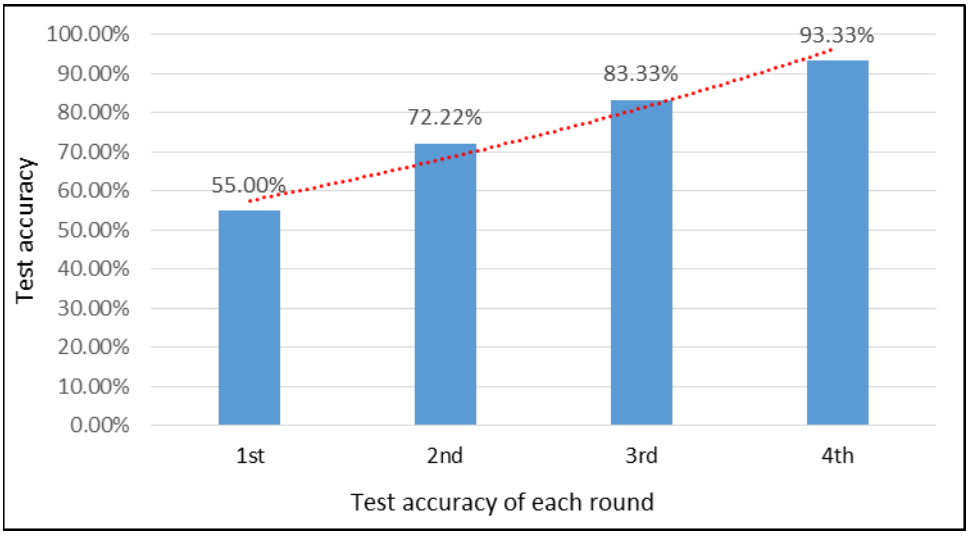

Figure 5.4: The change of test accuracy on the Scaled dataset

5.3 Summary

In this chapter, we examine the characteristics of the LAD technique, including the applicability of inventory classification and the capability of detecting inconsistencies in classification. The tests on two numerical examples have demonstrated that LAD is not only capable of classifying inventory, but also of detecting and correcting inconsistent observations when combined with the RCA procedure.

CHAPTER 6 COMPARISON WITH BENCHMARK TECHNIQUES

To illustrate the capability of the LAD technique for classification and its effectiveness in correcting inconsistencies in datasets, we compare the LAD classification results with other intelligence-based benchmark techniques, namely artificial neural network, support vector machines, k-nearest neighbour and Naïve Bayes. The four benchmark machine learning techniques have already been introduced in Chapter 2, Literature Review. The datasets used are the same as the two numerical examples in Chapter 5, LAD Classification on Numerical Examples. The classification results of the four machine learning classification techniques are compared with the LAD classification results in order to study the applicability and effectiveness of the LAD technique.

6.1 Introduction

This comparison includes two facets. One is a comparison between the original dataset and our new dataset of the same machine learning technique. Our new dataset here refers to the dataset after being corrected via LAD classification and the RCA procedure. Another aspect is a comparison among the five machine learning techniques for the same dataset.

6.1.1 Performance metrics

Supervised machine learning techniques have been widely used in classification problems. The performance of machine learning techniques is one of the most important aspects for selecting a classification algorithm. Although there is no empirical evaluation of supervised machine learning techniques, test accuracy (also known as predictive accuracy) is considered as a popular metric to evaluate the classification performance. In our study, we use test accuracy to make a comparison among the classification techniques. Test accuracy simply evaluates how often the classification algorithm makes the correct prediction. It is the ratio between the number of correct predictions and the total number of test observations. The formula is shown below.

Test accuracy = No. of correct predictions/ Total No. of test observations

6.1.2 Tools and configuration

There are four benchmark machine learning techniques, namely, the artificial neural network, support vector machines, k-nearest neighbour and Naïve Bayes. They are implemented by the software Weka which was developed by the Machine Learning Group at the University of Waikato, New Zealand (Hall et al., 2009). Weka, an open source software, is a collection of machine learning techniques and algorithms. Artificial neural network, k-nearest neighbour and Naïve Bayes are integrated in the software, and the package of support vector machines is installed from a third-party wrapper called 'LIBSVM'.

In our study, the Gaussian Radial Basis Function (RBF) kernel is used for SVM. The RBF kernel has the following form:

$K(x, x') = \exp(-\gamma \|x - x'\|^2)$

C is the regularization parameter and γ is the interval (width) of the RBF function. The best (C, γ) parameters can be determined by a grid search within all grid points of (C, γ) with the highest cross validation accuracy. Then the best parameters are applied to train the training data and the classifier model is generated (C.-C. Chang & Lin, 2011).

The Naive Bayes classifier provides a simple and efficient way to solve a classification problem by using an estimator for classes in supervised tasks. The Naive Bayes traditionally has the assumption that numeric attributes conform to a Gaussian distribution. But this is not always the case in the real world. John & Langley (1995) suggested that kernel estimation is a useful tool for building Bayesian models. In our study, we use two estimators to classify observations and choose the best cross validation accuracy for the final results.

K-Nearest Neighbours (KNN) is a non-parametric classification technique. The principle of KNN is to compute the distance or similarity measure between neighbor observations. We use Euclidean distance as a measure of distance or a similarity, which is one of the most popular measures. The selection of the K value is essential for applying KNN. By choosing the parameter of K from 1 to 5 in our study, as recommended by Malhotra, Sharma, & Nair (1999), we get the best cross validation accuracy for the final results.

The Gaussian Radial Basis Function (RBF) is used as an activation function to train the hidden layer of the networks. The loss function is determined by the penalised squared error along with a

quadratic penalty on the non-bias weights in the output layer, and the parameters of the network can be solved through the loss function (Frank, 2014).

6.2 Comparison of spare parts inventory

The spare parts inventory presents three datasets (CRS, VRS and AHP), used as numerical examples from Chapter 5. Each dataset evolves into a new dataset after the LAD classification and the RCA procedure. The new datasets are marked CRS-N, VRS-N and AHP-N.

Artificial neural network (RBF-Classifier), SVM, KNN and Naïve Bayes are implemented by WEKA. The LAD training and testing is implemented by cbmLAD. The test accuracy of the classification result is compared between the original and new datasets. The ten-fold cross validation is applied to each machine learning technique and test accuracy is measured by the percentage of correctly classified observations.

6.2.1 Comparison between original and new datasets

We get the test accuracy of original datasets and new datasets after a ten-fold cross validation. The result shows that new datasets have a higher test accuracy in each machine learning technique than we expected. The detail of the test accuracy is shown in Table 6.1. The numbers in red represent the best test accuracy among machine learning techniques on same dataset.

Table 6.1: Test accuracy on original and new datasets

Machine learning techniques	Datasets					
	CRS	CRS-N	VRS	VRS-N	AHP	AHP-N
SVM	55%	70%	65%	85%	85%	100%
Naïve Bayes	55%	100%	65%	100%	85%	90%
KNN	65%	75%	60%	75%	85%	95%
RBF-Classifier	45%	75%	60%	85%	90%	95%
LAD	60%	85%	45%	100%	85%	90%

The comparison of test accuracy for the original datasets and new datasets over five classification techniques is shown in Figure 6.1.

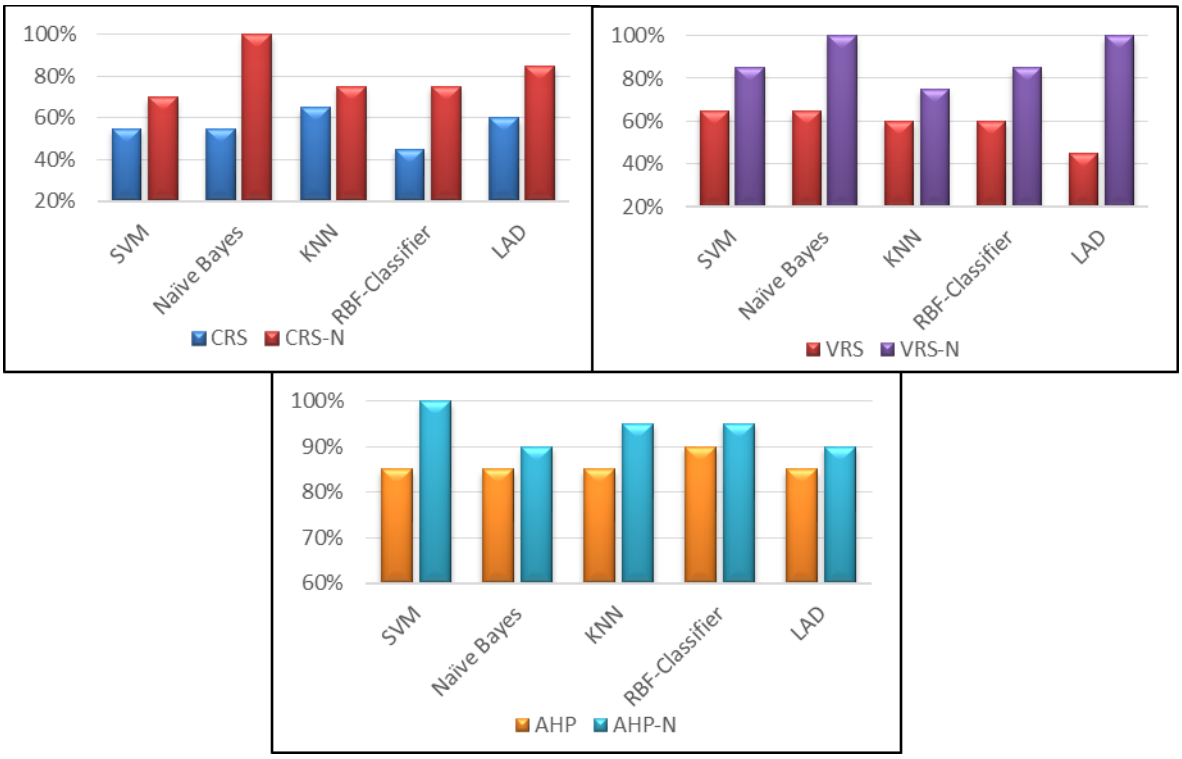

Figure 6.1: Comparison of test accuracy between original datasets and new datasets

It can be observed that test accuracy on 'VRS-N' has the best improvement, 55%, by LAD. The most improvement on 'CRS-N' is 45%, by Naïve Bayes. The counterpart on 'AHP-N' is 15% by SVM. Table 6.2 shows details about the improvement in test accuracy on new datasets. The numbers in red represent the most improvement in test accuracy among machine learning techniques on the same dataset.

Table 6.2: Test accuracy improvement on new datasets

Machine learning techniques	Improvement on new datasets		
	CRS-N	VRS-N	AHP-N
SVM	15%	20%	15%
Naïve Bayes	45%	35%	5%
KNN	10%	15%	10%
RBF-Classifier	30%	25%	5%
LAD	25%	55%	5%

6.2.2 Comparison among classification of machine learning techniques

As for different classification techniques over all of the original datasets, KNN has the best average accuracy, with 70%. SVM and Naïve Bayes techniques share second place, with 68% test accuracy. RBF-Classifier and the LAD technique rank third with 65% and fourth with 63%, respectively.

For new datasets, we find that the average accuracy of all learning techniques improves considerably. The Naïve Bayes technique reaches the best average test accuracy with 97%. The LAD technique ranks second with 92%. The KNN technique has a relatively small improvement in average accuracy, from 70% to 82%. SVM and RBF-Classifier gain improvements to 85%.

The details of test accuracy of classification of machine learning techniques on both original datasets and new datasets are shown in Table 6.3. The numbers in red represent the best test accuracy among machine learning techniques on the same dataset. The comparisons of test accuracy of each machine learning technique for both original datasets and new datasets are shown in Figure 6.2.

Table 6.3: Test accuracy by machine learning techniques

Datasets	Machine learning techniques				
	SVM	Naïve Bayes	KNN	RBF-Classifier	LAD
CRS	55%	55%	65%	45%	60%
VRS	65%	65%	60%	60%	45%
AHP	85%	85%	85%	90%	85%
Average	68%	68%	70%	65%	63%
CRS-N	70%	100%	75%	75%	85%
VRS-N	85%	100%	75%	85%	100%
AHP-N	100%	90%	95%	95%	90%
Average	85%	97%	82%	85%	92%

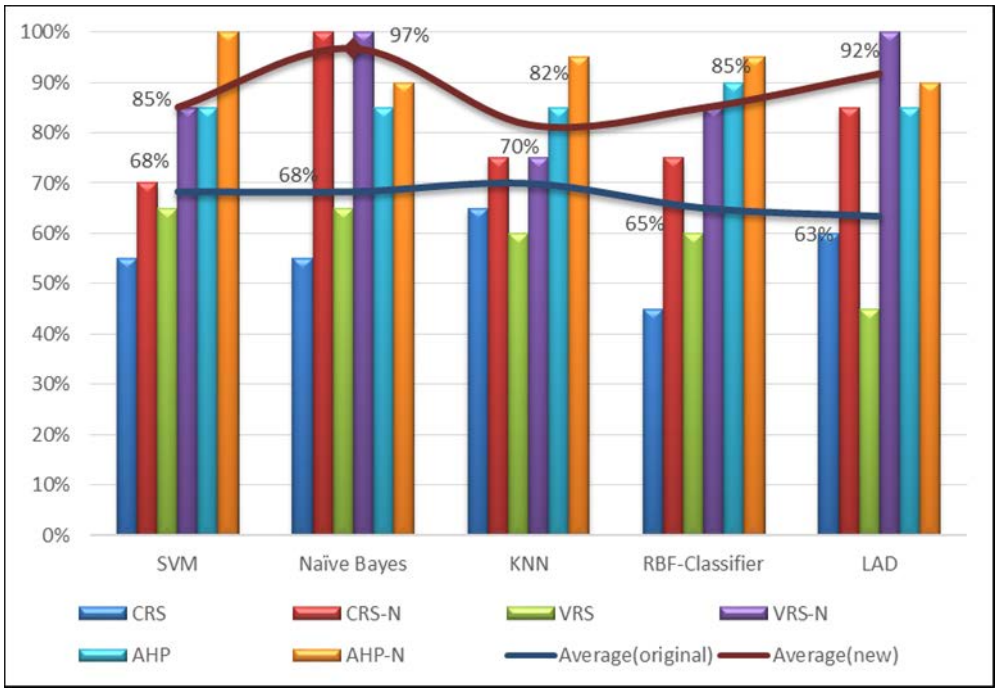

Figure 6.2: Comparison on test accuracy improvement of machine learning techniques

LAD has the most improvement, 55% on 'VRS-N' and Naïve Bayes receives the best improvement, 45% on 'CRS-N'. SVM has the highest increase, 15% on the 'AHP-N'. Table 6.4 shows the detail of improvement in test accuracy on new datasets in five machine learning techniques. The numbers in red represent the most improvement of test accuracy among machine learning techniques on the same dataset.

Table 6.4: Test accuracy improvement by machine learning techniques

Datasets	Test accuracy improvement by machine learning techniques				
	SVM	Naïve Bayes	KNN	RBF-Classifier	LAD
CRS-N	15%	45%	10%	30%	25%
VRS-N	20%	35%	15%	25%	55%
AHP-N	15%	5%	10%	5%	5%

6.3 Comparison of medical equipment inventory

Medical equipment spare parts inventory has been used by a number of researchers for the demonstration of classification techniques (Hadi-Vencheh, 2010; Ng, 2007; Soylu & Akyol, 2014; Yu, 2011). The inventory presents three datasets: AHP, Optimal, and Scaled three datasets. More information on this inventory can be found in Chapter 5. The three datasets evolve into new datasets

after the LAD classification and the RCA procedure. The new datasets are marked AHP-N, Optimal-N and Scaled-N. Artificial neural network (RBF-Classifier), SVM, KNN and Naïve Bayes are implemented by WEKA and LAD learning and testing is implemented by cbmLAD.

6.3.1 Comparison between original and new datasets

We achieve test accuracy by using ten-fold cross validation. Test accuracy is measured by the percentage of correctly classified observations. In the original datasets, the best accuracy is 83% on 'Scaled' by KNN technique. In the new datasets, the best accuracy is 95.74% on 'Scaled-N' and 'Optimal-N' both by LAD. Table 6.5 shows the detail of test accuracy on the original and new datasets. The numbers in red represent the best test accuracy among machine learning techniques on the same dataset.

Table 6.5: Test accuracy on original and new datasets

Machine learning techniques	Datasets					
	Scaled	Scaled-N	Optimal	Optimal-N	AHP	AHP-N
SVM	61.70%	85.11%	48.94%	80.85%	57.45%	80.85%
Naïve Bayes	61.70%	89.36%	59.57%	89.36%	59.57%	85.11%
KNN	83.00%	93.62%	42.55%	85.12%	63.83%	72.34%
RBF-Classifier	76.60%	93.62%	57.45%	93.62%	72.34%	93.62%
LAD	70.21%	95.74%	42.55%	95.74%	70.21%	85.11%

The comparisons of test accuracy between original datasets and new datasets over five machine learning techniques are shown in Figure 6.3.

Table 6.6: Test accuracy improvement on new datasets

Machine learning techniques	Improvement on new datasets		
	Scaled-N	Optimal-N	AHP-N
SVM	23%	32%	23%
Naïve Bayes	28%	30%	26%
KNN	11%	43%	9%
RBF-Classifier	17%	36%	21%
LAD	26%	53%	15%

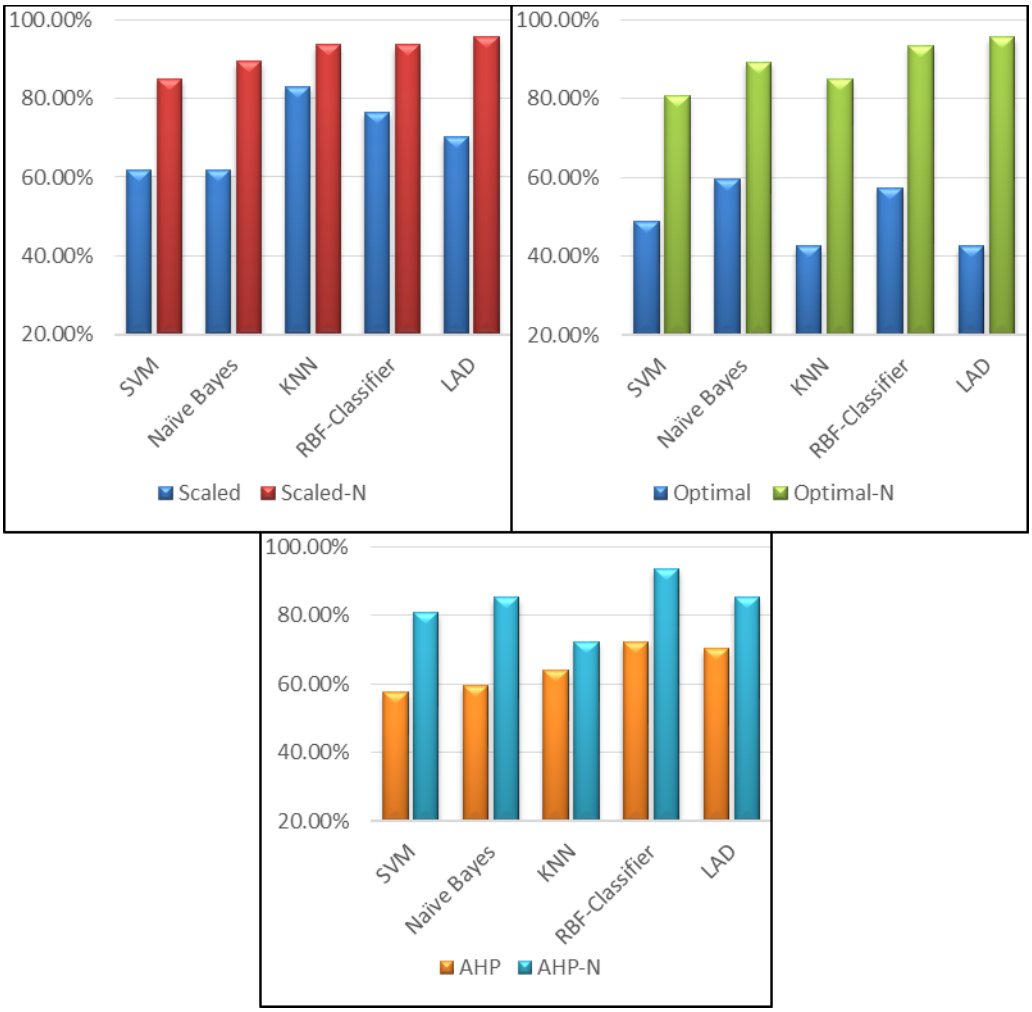

Figure 6.3: Comparison of test accuracy between original datasets and new datasets

We can see that test accuracy on 'Optimal-N' receives the best improvement, 53%, by LAD. The most improvement on 'Scaled-N' is 28%, by Naïve Bayes. The counterpart on 'AHP-N' is 26% by Naïve Bayes as well. Table 6.6 show the detail of test accuracy improvement on new datasets. The numbers in red represent the most improvement of test accuracy among machine learning techniques on the same dataset.

6.3.2 Comparison among classification of machine learning techniques

As for classification of machine learning techniques in all of the original datasets, RBF-Classifier reaches the best average accuracy, with 69%. KNN takes second place with 63% accuracy. The LAD technique ranks third with 61%. Naïve Bayes and SVM techniques have 60% and 56% accuracy, respectively. The difference in test accuracy among these classification techniques is not

profound. The detail of test accuracy by classification techniques on both original and new datasets is shown in Table 6.7.

On new datasets, we notice that the average test accuracy of all machine learning techniques is improving. RBF-Classifier technique reaches the best average test accuracy with 94%. The LAD technique ranks second with 92%. Naïve Bayes takes third place with 88%. KNN and SVM techniques receive 84% and 82%, respectively. Figure 6.4 shows the detail in test accuracy of each machine learning technique. The numbers in red represent the best test accuracy among machine learning techniques on the same dataset.

Table 6.7: Test accuracy by machine learning technique

Datasets	Machine learning techniques				
	SVM	Naïve Bayes	KNN	RBF-Classifier	LAD
Scaled	61.70%	61.70%	83.00%	76.60%	70.21%
Optimal	48.94%	59.57%	42.55%	57.45%	42.55%
AHP	57.45%	59.57%	63.83%	72.34%	70.21%
Average	56%	60%	63%	69%	61%
Scaled-N	85.11%	89.36%	93.62%	93.62%	95.74%
Optimal-N	80.85%	89.36%	85.12%	93.62%	95.74%
AHP-N	80.85%	85.11%	72.34%	93.62%	85.11%
Average	82%	88%	84%	94%	92%

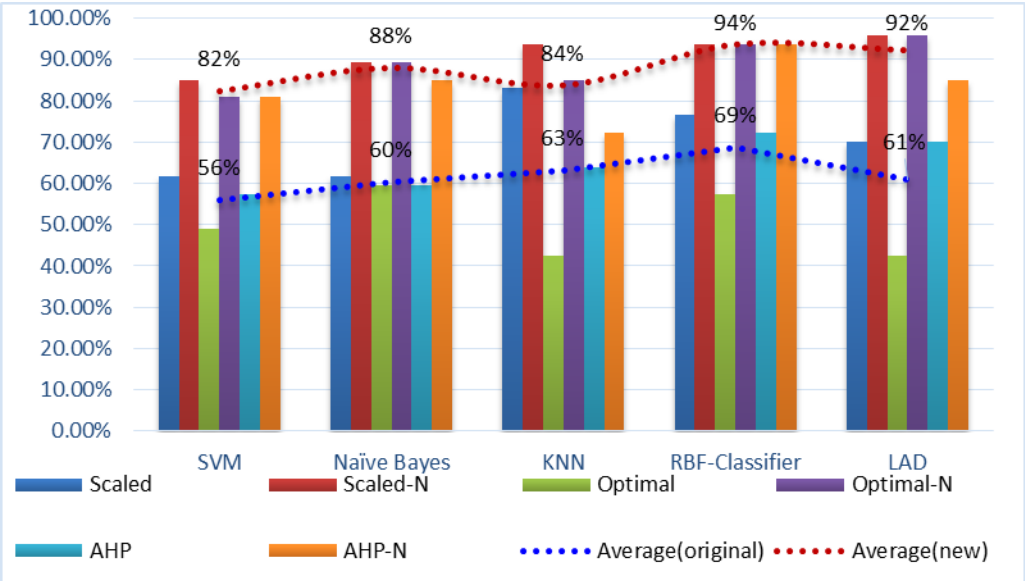

Figure 6.4: Comparison of test accuracy improvement of machine learning techniques

LAD enjoys the most improvement, 53% on ‘Optimal-N’ and Naïve Bayes receives the best improvement, 28% on ‘Scaled-N’ and 26% on ‘AHP-N’. The best average accuracy improvement is 31% by LAD. Table 6.8 shows the detail of improvement in test accuracy on the new datasets through five machine learning techniques.

Table 6.8: Test accuracy improvement by machine learning technique

Datasets	Test accuracy improvement by learning technique				
	SVM	Naïve Bayes	KNN	RBF-Classifier	LAD
Scaled-N	23%	28%	11%	17%	26%
Optimal-N	32%	30%	43%	36%	53%
AHP-N	23%	26%	9%	21%	15%

6.4 Statistical analysis

In previous sections, we have compared test accuracy on the original and new datasets regardless of machine learning technique. We also studied the test accuracy of five machine learning classification techniques. The figures in Section 6.2 and 6.3 illustrate the improvement in new datasets and machine learning techniques. In this section, we implement a statistical analysis to evaluate the differences between original and new datasets and the differences among five machine learning techniques.

6.4.1 Statistical analysis between the original and new datasets

In previous sections, the comparisons of spare parts inventory and medical equipment inventory have revealed the improvement in test accuracy after LAD classification and the RCA procedure. Each dataset evolves into a new dataset after the LAD classification and the RCA procedure. In order to study the difference in test accuracy between the original and new datasets, a paired T-test is carried out to investigate the hypotheses in statistical perspective.

6.4.1.1 Paired T-test for datasets of spare parts inventory

The spare parts inventory datasets include three original datasets marked as CRS, VRS and AHP, respectively. The new datasets are marked as CRS-N, VRS-N and AHP-N after the correction of inconsistencies. The test accuracy on the original and new datasets through five machine learning classification techniques is displayed in Table 6.9. The symbol of ‘↑’ stands for the improvement

in test accuracy. The numbers in red represent the best test accuracy among the machine learning techniques on the same dataset.

Table 6.9: Test accuracy on original and new datasets

Machine learning techniques	Datasets								
	CRS	CRS-N	↑	VRS	VRS-N	↑	AHP	AHP-N	↑
SVM	55%	70%	15%	65%	85%	20%	85%	100%	15%
Naïve Bayes	55%	100%	45%	65%	100%	35%	85%	90%	5%
KNN	65%	75%	10%	60%	75%	15%	85%	95%	10%
RBF-Classifier	45%	75%	30%	60%	85%	25%	90%	95%	5%
LAD	60%	85%	25%	45%	100%	55%	85%	90%	5%

We use Microsoft Excel to perform the paired t-Tests. The significance level $\alpha = 0.01$ and one-tailed tests are chosen for the hypothesis tests. We expect that new datasets have better performance in test accuracy regardless of machine learning technique. The paired t-Tests are implemented between 'CRS' and 'CRS-N', between 'VRS' and 'VRS-N' and between 'AHP' and 'AHP-N'. The results are shown in Table 6.10-6.12.

The relevant one-tailed hypotheses for the examination would be:

H1: Mean differences are greater than zero, $\mu_{diff} > 0$

H0: Mean differences are zero, $\mu_{diff} = 0$

Table 6.10: Paired t-Test for CRS and CRS-N

t-Test: Paired Two Sample for Means ($\alpha = 0.01$)	*CRS*	*CRS-N*
Mean	0.56	0.81
Observations	5	
Pearson Correlation	0.056478249	
H0: Null hypothesis	$\mu_{diff} = 0$	
H1: Alternative hypothesis	$\mu_{diff} > 0$	
df	4	
t Stat	-4.082482905	
P(T<=t) one-tail	0.007533849	
t Critical one-tail	3.746947388	
Decision	**Reject H0 and accept H1**	

Table 6.11: Paired t-Test for VRS and VRS-N

t-Test: Paired Two Sample for Means (α = 0.01)	*VRS*	*VRS-N*
Mean	0.59	0.89
Observations	5	
Pearson Correlation	-0.364932499	
H0: Null hypothesis	$\mu_{diff} = 0$	
H1: Alternative hypothesis	$\mu_{diff} > 0$	
df	4	
t Stat	-4.242640687	
P(T<=t) one-tail	0.0066178	
t Critical one-tail	3.746947388	
Decision	**Reject H0 and accept H1**	

Table 6.12: Paired t-Test for AHP and AHP-N

t-Test: Paired Two Sample for Means (α = 0.01)	*AHP*	*AHP-N*
Mean	0.86	0.94
Observations	5	
Pearson Correlation	0.133630621	
H0: Null hypothesis	$\mu_{diff} = 0$	
H1: Alternative hypothesis	$\mu_{diff} > 0$	
df	4	
t Stat	-4	
P(T<=t) one-tail	0.008065	
t Critical one-tail	3.7469474	
Decision	**Reject H0 and accept H1**	

We find that all p-values are less than α (0.01); meanwhile, all t Stats (absolute value) are larger than the values of t Critical one-tail. Both p-value and t -statistic methods reject the null hypothesis H0 and accept the alternate hypothesis H1. Thus, we can say that there is statistical significance at level of α (0.01). In other words, the new datasets have better performance of test accuracy at a confidence level of 99%, regardless of which machine learning techniques are used.

6.4.1.2 Paired T-test for datasets of medical equipment inventory

The medical equipment inventory datasets, classified by three classification methods, are marked as Scaled, Optimal and AHP. The new datasets are correspondingly marked as Scaled-N, Optimal-

N and AHP-N after the correction of inconsistencies. Test accuracy of the five machine learning classification techniques is displayed in Table 6.13. The numbers in red represent the best test accuracy among machine learning techniques on the same dataset.

Table 6.13: Test accuracy on original and new datasets

Machine learning techniques	Datasets					
	Scaled	Scaled-N	Optimal	Optimal-N	AHP	AHP-N
SVM	61.70%	85.11%	48.94%	80.85%	57.45%	80.85%
Naïve Bayes	61.70%	89.36%	59.57%	89.36%	59.57%	85.11%
KNN	83.00%	93.62%	42.55%	85.12%	63.83%	72.34%
RBF-Classifier	76.60%	93.62%	57.45%	93.62%	72.34%	93.62%
LAD	70.21%	95.74%	42.55%	95.74%	70.21%	85.11%

Microsoft Excel is used to perform the paired t-Tests. The improvement in test accuracy appears to be more obvious here compared to spare parts inventory. So, we choose the significance level $\alpha = 0.005$ and one-tailed tests for the hypothesis tests. We presume that new datasets have better performance in test accuracy regardless of machine learning technique. The paired t-Tests are implemented between Scaled and Scaled-N datasets, between Optimal and Optimal-N datasets and between AHP and AHP-N datasets. The results are shown in Table 6.14-6.16.

The relevant one-tailed hypotheses for the examination would be:

H1: Mean differences are greater than zero, $\mu_{diff} > 0$

H0: Mean differences are zero, $\mu_{diff} = 0$

Table 6.14: Paired t-Test for Scaled and Scaled-N

t-Test: Paired Two Sample for Means ($\alpha = 0.005$)	*Scaled*	*Scaled-N*
Mean	0.70642	0.9149
Observations	5	
Pearson Correlation	0.71340308	
H0: Null hypothesis	$\mu_{diff} = 0$	
H1: Alternative hypothesis	$\mu_{diff} > 0$	
df	4	
t Stat	-6.691236177	
P(T<=t) one-tail	0.001297316	
t Critical one-tail	4.604094871	
Decision	**Reject H0 and accept H1**	

Table 6.15: Paired t-Test for Optimal and Optimal-N

t-Test: Paired Two Sample for Means (α = 0.005)	*Optimal*	*Optimal-N*
Mean	0.50212	0.88938
Observations	5	
Pearson Correlation	0.128874332	
H0: Null hypothesis	$\mu_{diff} = 0$	
H1: Alternative hypothesis	$\mu_{diff} > 0$	
df	4	
t Stat	-9.168302943	
P(T<=t) one-tail	0.000392901	
t Critical one-tail	4.604094871	
Decision	**Reject H0 and accept H1**	

Table 6.16: Paired t-Test for AHP and AHP-N

t-Test: Paired Two Sample for Means (α = 0.005)	*AHP*	*AHP-N*
Mean	0.6468	0.83406
Observations	5	
Pearson Correlation	0.53247924	
H0: Null hypothesis	$\mu_{diff} = 0$	
H1: Alternative hypothesis	$\mu_{diff} > 0$	
df	4	
t Stat	-6.015584643	
P(T<=t) one-tail	0.001922895	
t Critical one-tail	4.604094871	
Decision	**Reject H0 and accept H1**	

Not only do we find that all p-values are less than α (0.005), but also all t Stats (absolute value) are larger than the values of t Critical one-tail. Both p-value and t -statistic methods reject the null hypothesis H0 and accept the alternate hypothesis H1. So there is statistical significance at the level of α (0.005). In other words, the new datasets have better performance in test accuracy regardless of which machine learning technique is applied at a confidence level of 99.5%.

6.4.2 Statistical analysis between datasets and learning techniques

In a previous section, we used a statistical analysis of paired t-Test to demonstrate the significant improvement in test accuracy on datasets after the correction of inconsistences by the LAD

technique and RCA procedure. We are also interested in the performance difference among learning techniques. To find out what learning techniques have better performance, the Friedman test is carried out. The procedure for the Friedman test consists of the following steps:

1. Collect data in the form of a matrix which has n rows (blocks) and k columns (treatments). Here, blocks are datasets and treatments are learning techniques.
2. Rank the data entry within each dataset from smallest to largest and transform the value to 1 (smallest) to k (biggest); if there are tied values, assign to each tied value with the average of the ranks (see example of Figure 6.5);
3. Define the null hypothesis H0 and the alternative hypothesis H1;
4. Calculate the Friedman test statistic $F_R = \frac{12}{nk(k+1)} \sum R^2 - 3n(k+1)$, n is the blocks of raters, k is the number of treatments and R is the sum of ranks for each treatment;
5. Select the significance level α and find the Chi-Squared value in terms of α and degrees of freedom k-1;
6. State the decision: Reject H0 if $F_R > \chi^2_\alpha$; otherwise, do not reject H0.

datasets (blocks)	Machine learning techniques (treatments)				
	SVM	Naïve Bayes	KNN	RBF-Classifier	LAD
CRS	55.00%	55.00%	65.00%	45.00%	60.00%
VRS	65.00%	65.00%	60.00%	60.00%	45.00%
AHP	85.00%	85.00%	85.00%	90.00%	85.00%

1: Numbers are arranged in ascending order for each block.

2: If there are tied values, assign to each tied value with the average of the ranks

CRS	rank (with ties)	rank
45.00%	1	1
55.00%	2	2.5
55.00%	3	2.5
60.00%	4	4
65.00%	5	5

AHP	rank (with ties)	rank
85.00%	1	2.5
85.00%	2	2.5
85.00%	3	2.5
85.00%	4	2.5
90.00%	5	5

VRS	rank (with ties)	rank
45.00%	1	1
60.00%	2	2.5
60.00%	3	2.5
65.00%	4	4.5
65.00%	5	4.5

datasets (blocks)	Machine learning techniques (treatments)									
	SVM	Rank	Naïve Bayes	Rank	KNN	Rank	RBF-Classifier	Rank	LAD	Rank
CRS	55.00%	2.5	55.00%	2.5	65.00%	5	45.00%	1	60.00%	4
VRS	65.00%	4.5	65.00%	4.5	60.00%	2.5	60.00%	2.5	45.00%	1
AHP	85.00%	2.5	85.00%	2.5	85.00%	2.5	90.00%	5	85.00%	2.5

Figure 6.5: Example of ranking data

6.4.2.1 Friedman Test: Nonparametric analysis for machine learning techniques

We investigate the performance difference among five machine learning techniques. The test accuracy results regarding machine learning techniques with ranks are shown in Table 6.17 and 6.18 for spare parts inventory (1st inventory) and Table 6.19 and 6.20 for medical equipment inventory (2nd inventory), respectively. In the next step, we conduct the Friedman test by following the procedure described above. The hypotheses would be as follows:

H0: The machine learning techniques have identical test accuracy;

H1: Not all machine learning techniques have identical test accuracy.

Table 6.17: Test accuracy of machine learning techniques on original datasets (1st inventory)

Datasets	Machine learning techniques									
	SVM		Naïve Bayes		KNN		RBF-Classifier		LAD	
	Test Accuracy	Rank	Test Accuracy	Rank	Test Accuracy	Rank	Test Accuracy	Rank	Test Accuracy	Rank
CRS	55.00%	2.5	55.00%	2.5	65.00%	5	45.00%	1	60.00%	4
VRS	65.00%	4.5	65.00%	4.5	60.00%	2.5	60.00%	2.5	45.00%	1
AHP	85.00%	2.5	85.00%	2.5	85.00%	2.5	90.00%	5	85.00%	2.5
Rank Total(R)		9.5		9.5		10		8.5		7.5

Table 6.18: Test accuracy of machine learning techniques on new datasets (1st inventory)

Datasets	Machine learning techniques									
	SVM		Naïve Bayes		KNN		RBF-Classifier		LAD	
	Test Accuracy	Rank	Test Accuracy	Rank	Test Accuracy	Rank	Test Accuracy	Rank	Test Accuracy	Rank
CRS-N	70.00%	1	100.00%	5	75.00%	2.5	75.00%	2.5	85.00%	4
VRS-N	85.00%	2.5	100.00%	4.5	75.00%	1	85.00%	2.5	100.00%	4.5
AHP-N	100.00%	5	90.00%	1.5	95.00%	3.5	95.00%	3.5	90.00%	1.5
Rank Total(R)		8.5		11		7		8.5		10

Table 6.19: Test accuracy of machine learning techniques on original datasets (2nd inventory)

Datasets	Machine learning techniques									
	SVM		Naïve Bayes		KNN		RBF-Classifier		LAD	
	Test Accuracy	Rank	Test Accuracy	Rank	Test Accuracy	Rank	Test Accuracy	Rank	Test Accuracy	Rank
Scaled	61.70%	1.5	61.70%	1.5	83.00%	5	76.60%	4	70.21%	3
Optimal	48.94%	3	59.57%	5	42.55%	1.5	57.45%	4	42.55%	1.5
AHP	57.45%	1	59.57%	2	63.83%	3	72.34%	5	70.21%	4
Rank Total(R)		5.5		8.5		9.5		13		8.5

Table 6.20: Test accuracy of machine learning techniques on new datasets (2nd inventory)

Datasets	Machine learning techniques									
	SVM		Naïve Bayes		KNN		RBF-Classifier		LAD	
	Test Accuracy	Rank	Test Accuracy	Rank	Test Accuracy	Rank	Test Accuracy	Rank	Test Accuracy	Rank
Scaled-N	85.11%	1	89.36%	2	93.62%	3.5	93.62%	3.5	95.74%	5
Optimal-N	80.85%	1	89.36%	3	85.12%	2	93.62%	4	95.74%	5
AHP-N	80.85%	2	85.11%	3.5	72.34%	1	93.62%	5	85.11%	3.5
Rank Total(R)		4		8.5		6.5		12.5		13.5

The results of the Friedman test are shown in Table 6.21.

Table 6.21: The results of the Friedman test for machine learning techniques

Source	Table 6.17	Table 6.18	Table 6.19	Table 6.20
n	3	3	3	3
k	5	5	5	5
$\sum R^2$	409	414.5	434	469
α	**0.05**	**0.05**	**0.05**	**0.05**
χ^2_α	9.488	9.488	9.488	9.488
Friedman test statistic F_R	**0.5333**	**1.2667**	**3.8667**	**8.5333**
p value	0.9702	0.8670	0.4243	0.0739
Decision	**Accept H0**	**Accept H0**	**Accept H0**	**Accept H0**

From Table 6.21, we find that that the values of the Friedman test statistic F_R are less than χ^2_α on original and new datasets of both inventories, which means we accept H0. The conclusion is that there is insufficient evidence to support the significant difference in test accuracy among the five machine learning techniques.

6.5 Summary

In this chapter, we have employed five machine learning techniques, namely artificial neural network (RBF-Classifier), SVM, KNN, Naïve Bayes and LAD, to classify the original and new datasets of both spare parts inventory and medical equipment inventory. The test accuracy with ten-fold cross validation is evaluated for each classification technique and each dataset.

We compare the results in test accuracy of the original and new datasets in two ways. First, the test accuracy of the original and new datasets is examined. The new datasets have better performance on all machine learning techniques. We also conduct the Paired t-Test to prove the significant difference between the original and new datasets of both inventories. Secondly, the test accuracy performance on datasets and machine learning techniques is investigated. The results of the Friedman Test show that there is insufficient evidence to support the significant difference in test accuracy of five machine learning techniques either in the original or the new datasets of both inventories.

CHAPTER 7 CONCLUSION AND FUTURE WORK

7.1 Conclusion

The overall objective of this research is to study the applicability of inventory classification and the capability for correcting classification inconsistencies through the LAD technique. Inventory classification has attracted attention from many researchers. Techniques of inventory classification are well studied, from traditional ABC analysis (single criterion) to multi-criteria classification and from mathematical formulation to machine learning algorithms. Although the traditional ABC analysis has a reputation for ease of use, it cannot meet the classification requirements of sophisticated products. Mathematical models and methods would raise complexity and computing costs increase dramatically when many criteria are taken into account for classification. Machine learning techniques are more suitable for multi-criteria classification and provide more accurate classification. The LAD technique, as a machine learning approach, provides a new ability for inventory classification.

Moreover, the LAD technique is interpretable and transparent for classification. Patterns are generated in the training stage and they can be used to explain the results of classification. Benefiting from the characteristics of the LAD technique and RCA procedure, we are able to find and fix inconsistent observations.

We use two examples to demonstrate LAD's capacities for inventory classification and inconsistency corrections. The results show that the test accuracy improved significantly for new datasets by both LAD and other machine learning classification techniques. The statistical analysis is applied to confirm the significant difference between the original and new (corrected) datasets. Among five machine learning techniques, namely artificial neural network, support vector machines, AHP, k-nearest neighbour, Naïve Bayes and LAD, the statistical analysis shows that there is insufficient evidence to support the significant difference of test accuracy either on the original datasets or in the new datasets.

In summary, we have reached three main conclusions:

1. We have demonstrated the LAD technique for making multi-criteria classifications for spare parts inventory;

2. We have investigated the LAD techniques that are capable of detecting and correcting inconsistencies along with RCA procedures and improved test accuracy;
3. We have provided evidence that LAD is a competitive technique for classification by comparing other machine learning classification techniques.

7.2 Future work

The capability of the LAD technique is very encouraging for inventory classification. The LAD classification relies on criteria that we collect, also known as attributes. The accuracy of attributes greatly impacts the classification results. For instance, the attribute ‘cost’ has a more accurate meaning in accounting than the attribute ‘price’ for classification. In the numerical examples, we simply use price as the cost of individual inventory. This is not the case of cost in the real world. The cost of inventory is not only affected by price, but also by currency exchange rates, expedited service fees, etc.

Another possible improvement practice for classification would be to introduce more attributes, such as demand forecasting. A multiple dimension product is determined by many characteristics. Machine learning techniques make it possible to classify inventory based on many different factors.

So far, we have studied the inventory classification in a supervised learning style. If we had a completely new dataset without classes, the unsupervised learning classification would be one way to classify the inventory. An investigation of the LAD technique’s capabilities in unsupervised learning would also be a good subject to study.

BIBLIOGRAPHY

Agarwal, S., Jain, N., & Dholay, S. (2015). Adaptive Testing and Performance Analysis Using Naive Bayes Classifier. *Procedia Computer Science, 45*, 70-75. doi:10.1016/j.procs.2015.03.088

Alexe, G., Alexe, S., Tib\, \#233, Bonates, r. O., & Kogan, A. (2007). Logical analysis of data --- the vision of Peter L. Hammer. *Annals of Mathematics and Artificial Intelligence, 49*(1-4), 265-312. doi:10.1007/s10472-007-9065-2

Alexe, G., & Hammer, P. L. (2006). Spanned patterns for the logical analysis of data. *Discrete Applied Mathematics, 154*(7), 1039-1049. doi:10.1016/j.dam.2005.03.031

Alexe, S., Blackstone, E., Hammer, P. L., Ishwaran, H., Lauer, M. S., & Pothier Snader, C. E. (2003). Coronary risk prediction by logical analysis of data. *Annals of Operations Research, 119*, 15-42. doi:10.1023/a:1022970120229

Altay Guvenir, H., & Erel, E. (1998). Multicriteria inventory classification using a genetic algorithm. *European Journal of Operational Research, 105*(1), 29-37. doi:http://dx.doi.org/10.1016/S0377-2217(97)00039-8

Avila-Herrera, J. F., & Subasi, M. M. (2015, 19-23 Oct. 2015). *Logical analysis of multi-class data.* Paper presented at the Computing Conference (CLEI), 2015 Latin American.

Babai, M. Z., Ladhari, T., & Lajili, I. (2015). On the inventory performance of multi-criteria classification methods: Empirical investigation. *International Journal of Production Research, 53*(1), 279-290. doi:10.1080/00207543.2014.952791

Bennane, A., & Yacout, S. (2012). LAD-CBM; New data processing tool for diagnosis and prognosis in condition-based maintenance. *Journal of Intelligent Manufacturing, 23*(2), 265-275. doi:10.1007/s10845-009-0349-8

Boros, E., Hammer, P. L., Ibaraki, T., Kogan, A., Mayoraz, E., & Muchnik, I. (2000). An Implementation of Logical Analysis of Data. *IEEE Transactions On Knowledge and Data Engineering, 12*(2), 292-306.

Boylan, J. E., & Syntetos, A. A. (2010). Spare parts management: a review of forecasting research and extensions. *Ima Journal of Management Mathematics, 21*, 227-237. doi:10.1093/imaman/dpp016

Braglia, M., Grassi, A., & Montanari, R. (2004). Multi-attribute classification method for spare parts inventory management. *Journal of Quality in Maintenance Engineering, 10*(1), 55-65. doi:10.1108/13552510410526875

Bramer, M. (2013). *Principles of Data Mining* (2 ed.): Springer London.

Cebi, F., Kahraman, C., & Bolat, B. (2010). *A multiattribute ABC classification model using fuzzy AHP*. Paper presented at the 2010 40th International Conference on Computers & Industrial Engineering (CIE-40 2010), 25-28 July 2010, Piscataway, NJ, USA.

Chang, C.-C., & Lin, C.-J. (2011). LIBSVM: A library for support vector machines. *ACM Trans. Intell. Syst. Technol., 2*(3), 1-27. doi:10.1145/1961189.1961199

Chang, F., Chou, C.-H., Lin, C.-C., & Chen, C.-J. (2004). *A prototype classification method and its application to handwritten character recognition.* Paper presented at the 2004 IEEE International Conference on Systems, Man and Cybernetics, SMC 2004, October 10, 2004 - October 13, 2004, The Hague, Netherlands.

Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research, 2*(6), 429-444. doi:10.1016/0377-2217(78)90138-8

Crama, Y., Hammer, P. L., & Ibaraki, T. (1988). Cause-effect relationships and partially defined Boolean functions. *Annals of Operations Research, 16*(1-4), 299-326.

Duan, K.-B., Rajapakse, J. C., & Nguyen, M. N. (2007). *One-versus-one and one-versus-all multiclass SVM-RFE for gene selection in cancer classification.* Paper presented at the 5th European Conference on Evolutionary Computation, Machine Learning and Data Mining in Bioinformatics, EvoBIO 2007, April 11, 2007 - April 13, 2007, Valencia, Spain.

Duchessi, P., Tayi, G. K., & Levy, J. B. (1988). A conceptual approach for managing of spare parts. *International Journal of Physical Distribution & Materials Management, 18*(5), 8-15. doi:10.1108/eb014700

Dupuis, C., Gamache, M., & Pagé, J.-F. (2012). Logical analysis of data for estimating passenger show rates at Air Canada. *Journal of Air Transport Management, 18*(1), 78-81. doi:http://dx.doi.org/10.1016/j.jairtraman.2011.10.004

Flores, B. E., Olson, D. L., & Dorai, V. K. (1992). Management of multicriteria inventory classification. *Mathematical and Computer Modelling, 16*(12), 71-82. doi:http://dx.doi.org/10.1016/0895-7177(92)90021-C

Foody, G. M., & Mathur, A. (2004). A relative evaluation of multiclass image classification by support vector machines. *IEEE Transactions on Geoscience and Remote Sensing, 42*(6), 1335-1343. doi:10.1109/TGRS.2004.827257

Frank, E. (2014) Fully Supervised Training of Gaussian Radial Basis Function Networks in WEKA. (pp. 1-4). Department of Computer Science, University of Waikato.

Fu, Y., Lai, K. K., Miao, Y., & Leung, J. W. K. (2015). A distance-based decision-making method to improve multiple criteria ABC inventory classification. doi:10.1111/itor.12193

Gajpal, P. P., Ganesh, L. S., & Rajendran, C. (1994). CRITICALITY ANALYSIS OF SPARE PARTS USING THE ANALYTIC HIERARCHY PROCESS. *International Journal of Production Economics, 35*(1-3), 293-297. doi:10.1016/0925-5273(94)90095-7

Galar, M., Fernandez, A., Barrenechea, E., Bustince, H., & Herrera, F. (2011). An overview of ensemble methods for binary classifiers in multi-class problems: Experimental study on one-vs-one and one-vs-all schemes. *Pattern Recognition, 44*(8), 1761-1776. doi:10.1016/j.patcog.2011.01.017

Gong, A., & Liu, Y. (2011). *Improved KNN classification algorithm by dynamic obtaining K.* Paper presented at the International Conference on Advanced Research on Electronic Commerce, Web Application, and Communication, ECWAC 2011, April 16, 2011 - April 17, 2011, Guangzhou, China.

Gongde, G., Hui, W., Bell, D., Yaxin, B., & Greer, K. (2003). *KNN model-based approach in classification.* Paper presented at the On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE. OTM Confederated International Conferences CoopIS, DOA, and ODBASE 2003. Proceedings, 3-7 Nov. 2003, Berlin, Germany.

Guosheng, H., & Guohong, Z. (2008). Comparison on neural networks and support vector machines in suppliers' selection. *Journal of Systems Engineering and Electronics, 19*(2), 316-320. doi:10.1016/S1004-4132(08)60085-7

Hadi-Vencheh, A. (2010). An improvement to multiple criteria ABC inventory classification. *European Journal of Operational Research, 201*(3), 962-965. doi:10.1016/j.ejor.2009.04.013

Hadi-Vencheh, A., & Mohamadghasemi, A. (2011). A fuzzy AHP-DEA approach for multiple criteria ABC inventory classification. *Expert Systems with Applications, 38*(4), 3346-3352. doi:10.1016/j.eswa.2010.08.119

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software: an update. *SIGKDD Explor. Newsl., 11*(1), 10-18. doi:10.1145/1656274.1656278

Hammer, P. L., & Bonates, T. O. (2006). Logical analysis of data - an overview: from combinatorial optimization to medical applications. *Annals of Operations Research, 148,* 203-225. doi:10.1007/s10479-006-0075-y

Hammer, P. L., Kogan, A., Simeone, B., & Szedmak, S. (2004). Pareto-optimal patterns in logical analysis of data. *Discrete Applied Mathematics, 144*(1-2), 79-102. doi:10.1016/j.dam.2003.08.013

Hatefi, S. M., Torabi, S. A., & Bagheri, P. (2014). Multi-criteria ABC inventory classification with mixed quantitative and qualitative criteria. *International Journal of Production Research, 52*(3), 776-786. doi:10.1080/00207543.2013.838328

Jiang, L., Wang, D., Cai, Z., & Yan, X. (2007). *Survey of improving Naive Bayes for classification.* Paper presented at the 3rd International Conference on Advanced Data Mining and Applications, ADMA 2007, August 6, 2007 - August 8, 2007, Harbin, China.

John, G. H., & Langley, P. (1995). *Estimating continuous distributions in Bayesian classifiers*. Paper presented at the Proceedings of the Eleventh conference on Uncertainty in artificial intelligence, Montreal, Quebec, Canada.

Joshi, A. J., Porikli, F., & Papanikolopoulos, N. P. (2012). Scalable Active Learning for Multiclass Image Classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 34*(11), 2259-2273. doi:10.1109/TPAMI.2012.21

Kabir, G., & Hasin, M. A. A. (2013). Multi-criteria inventory classification through integration of fuzzy analytic hierarchy process and artificial neural network. *International Journal of Industrial and Systems Engineering, 14*(1), 74-103. doi:10.1504/IJISE.2013.052922

Kalaivani, P., & Shunmuganathan, K. L. (2014). *An improved K-nearest-neighbor algorithm using genetic algorithm for sentiment classification.* Paper presented at the 2014 International Conference on Circuits, Power and Computing Technologies, ICCPCT 2014, March 20, 2014 - March 21, 2014, Nagercoil, Tamil Nadu, India.

Kennedy, W. J., Wayne Patterson, J., & Fredendall, L. D. (2002). An overview of recent literature on spare parts inventories. *International Journal of Production Economics, 76*(2), 201-215. doi:10.1016/s0925-5273(01)00174-8

Keskin, G. A., & Ozkan, C. (2013). Multiple Criteria ABC Analysis with FCM Clustering. *Journal of Industrial Engineering*, 827274 (827277). doi:10.1155/2013/827274

Lejeune, M. A., & Margot, F. (2011). Optimization for simulation: LAD accelerator. *Annals of Operations Research, 188*(1), 285-305. doi:10.1007/s10479-009-0518-3

Lewis, D. D., Yang, Y., Rose, T. G., & Li, F. (2004). RCV1: A new benchmark collection for text categorization research. *Journal of Machine Learning Research, 5*, 361-397.

Liu, J. S., Lu, L. Y. Y., Lu, W.-M., & Lin, B. J. Y. (2013). Data envelopment analysis 1978–2010: A citation-based literature survey. *Omega, 41*(1), 3-15. doi:http://dx.doi.org/10.1016/j.omega.2010.12.006

Lolli, F., Ishizaka, A., & Gamberini, R. (2014). New AHP-based approaches for multi-criteria inventory classification. *International Journal of Production Economics, 156*, 62-74. doi:10.1016/j.ijpe.2014.05.015

Malhotra, M. K., Sharma, S., & Nair, S. S. (1999). Decision making using multiple models. *European Journal of Operational Research, 114*(1), 1-14. doi:http://dx.doi.org/10.1016/S0377-2217(98)00037-X

Mejdoub, M., & Ben Amar, C. (2013). Classification improvement of local feature vectors over the KNN algorithm. *Multimedia Tools and Applications, 64*(1), 197-218. doi:10.1007/s11042-011-0900-4

Moreira, M. (2000). *The use of Boolean concepts in general classification contexts.* Ecole Polytechnique Federale de Lausanne. Retrieved from http://infoscience.epfl.ch/record/146177

Mortada, M.-A., Carroll Iii, T., Yacout, S., & Lakis, A. (2012). Rogue components: Their effect and control using logical analysis of data. *Journal of Intelligent Manufacturing, 23*(2), 289-302. doi:10.1007/s10845-009-0351-1

Mortada, M.-A., & Yacout, S. (2011). cbmLAD - Using logical analysis of data in condition based maintenance. *ICCRD2011 - 2011 3rd International Conference on Computer Research and Development, 4*, 30-34. doi:10.1109/iccrd.2011.5763847

Mortada, M.-A., Yacout, S., & Lakis, A. (2013). Fault diagnosis in power transformers using multi-class logical analysis of data. *Journal of Intelligent Manufacturing*, 1-11.

Murphy, K. P. (2012). *Machine Learning: A Probabilistic Perspective*. Cambridge, Massachusetts London, England: The MIT Press.

Nakamura, S., Markov, K., Nakaiwa, H., Kikui, G., Kawai, H., Jitsuhiro, T., . . . Yamamoto, S. (2006). The ATR multilingual speech-to-speech translation system. *IEEE Transactions on Audio, Speech and Language Processing, 14*(2), 365-375. doi:10.1109/TSA.2005.860774

Ng, W. L. (2007). A simple classifier for multiple criteria ABC analysis. *European Journal of Operational Research, 177*(1), 344-353. doi:http://dx.doi.org/10.1016/j.ejor.2005.11.018

Ou, G., Murphey, Y. L., & Lee, F. (2004). *Multiclass pattern classification using neural networks.* Paper presented at the Proceedings of the 17th International Conference on Pattern Recognition, ICPR 2004, August 23, 2004 - August 26, 2004, Cambridge, United kingdom.

Partovi, F. Y., & Anandarajan, M. (2002). Classifying inventory using an artificial neural network approach. *Computers & Industrial Engineering, 41*(4), 389-404. doi:http://dx.doi.org/10.1016/S0360-8352(01)00064-X

Partovi, F. Y., & Hopton, W. E. (1994). The analytic hierarchy process as applied to two types of inventory problems. *Production and Inventory Management Journal, 35*(1), 13-19.

Pendharkar, P. C. (2010). *A hybrid radial basis function and data envelopment analysis neural network for classification.*

Rad, T., Shanmugarajan, N., & Wahab, M. I. M. (2011). Classification of critical spares for aircraft maintenance. *2011 8th International Conference on Service Systems and Service Management (ICSSSM 2011),* 6. doi:10.1109/icsssm.2011.5959470

Ragab, A., Ouali, M.-S., Yacout, S., & Osman, H. (2014). Remaining useful life prediction using prognostic methodology based on logical analysis of data and Kaplan-Meier estimation. doi:10.1007/s10845-014-0926-3

Ramanathan, R. (2006). ABC inventory classification with multiple-criteria using weighted linear optimization. *Computers & Operations Research, 33*(3), 695-700. doi:dx.doi.org/10.1016/j.cor.2004.07.014

Ratanamahatana, C. A., & Gunopulos, D. (2003). Feature selection for the naive Bayesian classifier using decision trees. *Applied Artificial Intelligence, 17*(5-6), 475-487.

Reid, R. A. (1987). ABC METHOD IN HOSPITAL INVENTORY MANAGEMENT: A PRACTICAL APPROACH. *Production and Inventory Management Journal, 28*(4), 67-70.

Rezaei, J., & Dowlatshahi, S. (2010). A rule-based multi-criteria approach to inventory classification. *International Journal of Production Research, 48*(23), 7107-7126. doi:10.1080/00207540903348361

Rezaei, J., & Salimi, N. (2015). Optimal ABC inventory classification using interval programming. *International Journal of Systems Science, 46*(11), 1944-1952. doi:10.1080/00207721.2013.843215

Rifkin, R., Mukherjee, S., Tamayo, P., Ramaswamy, S., Yeang, C.-H., Angelo, M., . . . Mesirov, J. P. (2003). An analytical method for multiclass molecular cancer classification. *SIAM Review, 45*(4), 706-723. doi:10.1137/S0036144502411986

Rui, X., Anagnostopoulos, G. C., & Wunsch, D. C. I. I. (2007). Multiclass cancer classification using semisupervised ellipsoid ARTMAP and particle swarm optimization with gene expression data. *IEEE/ACM Transactions on Computational Biology and Bioinformatics, 4*(1), 65-77. doi:10.1109/TCBB.2007.1009

Ryoo, H. S., & Jang, I.-Y. (2009). MILP approach to pattern generation in logical analysis of data. *Discrete Applied Mathematics, 157*(4), 749-761. doi:10.1016/j.dam.2008.07.005

Salamanca, D. Y., Soumaya. (2007). *Condition based maintenance with Logical Analysis of Data*. Paper presented at the 7e Congrès international de génie industriel, Trois-Rivières, Canada.

Sarmah, S. P., & Moharana, U. C. (2015). Multi-criteria classification of spare parts inventories - A web based approach. *Journal of Quality in Maintenance Engineering, 21*(4), 456-477. doi:10.1108/JQME-04-2012-0017

Shamsaddini, R., Vesal, S. M., & Nawaser, K. (2015). A new model for inventory items classification through integration of ABC-Fuzzy and fuzzy analytic hierarchy process. *International Journal of Industrial and Systems Engineering, 19*(2), 239-261. doi:10.1504/IJISE.2015.067250

Simunovic, K., Simunovic, G., & Saric, T. (2009). Application of artificial neural networks to multiple criteria inventory classification. *Strojarstvo, 51*(4), 313-321.

Soylu, B., & Akyol, B. (2014). Multi-criteria inventory classification with reference items. *Computers and Industrial Engineering, 69*(1), 12-20. doi:10.1016/j.cie.2013.12.011

Srihari, S. N. (2000). Handwritten address interpretation: A task of many pattern recognition problems. *International Journal of Pattern Recognition and Artificial Intelligence, 14*(5), 663-674. doi:10.1142/S0218001400000441

Stoll, J., Kopf, R., Schneider, J., & Lanza, G. (2015). Criticality analysis of spare parts management: a multi-criteria classification regarding a cross-plant central warehouse strategy. *Production Engineering, 9*(2), 225-235. doi:10.1007/s11740-015-0602-2

Su, X.-Y., Zhou, X.-L., & Mo, Y. (2010). *Forecast of spare parts inventory risk level based on support vector machine.* Paper presented at the 17th International Conference on Industrial Engineering and Engineering Management, IE and EM2010, October 29, 2010 - October 31, 2010, Xiamen, China.

Tavassoli, M., Faramarzi, G. R., & Saen, R. F. (2014). Multi-criteria ABC inventory classification using DEA-discriminant analysis to predict group membership of new items. *International Journal of Applied Management Science, 6*(2), 171-189. doi:10.1504/IJAMS.2014.060904

Torabi, S. A., Hatefi, S. M., & Saleck Pay, B. (2012). ABC inventory classification in the presence of both quantitative and qualitative criteria. *Computers & Industrial Engineering, 63*(2), 530-537. doi:http://dx.doi.org/10.1016/j.cie.2012.04.011

van Kampen, T. J., Akkerman, R., & van Donk, D. P. (2012). SKU classification: a literature review and conceptual framework. *International Journal of Operations & Production Management, 32*(7), 850-876. doi:10.1108/01443571211250112

Wang, J.-C., Wang, J.-F., Lin, C.-B., Jian, K.-T., & Kuok, W.-H. (2006). *Content-based audio classification using support vector machines and independent component analysis.* Paper presented at the 18th International Conference on Pattern Recognition, ICPR 2006, August 20, 2006 - August 24, 2006, Hong Kong, China.

Webb, G. I., Boughton, J. R., & Wang, Z. (2005). Not so naive Bayes: Aggregating one-dependence estimators. *Machine Learning, 58*(1), 5-24. doi:10.1007/s10994-005-4258-6

Weizhu, C., Jun, Y., Benyu, Z., Zheng, C., & Qiang, Y. (2007). *Document transformation for multi-label feature selection in text categorization.* Paper presented at the 2007 7th IEEE International Conference on Data Mining (ICDM '07), 28-31 Oct. 2007, Piscataway, NJ, USA.

Yang, N., Muraleedharan, R., Kohl, J., Demirkol, I., Heinzelman, W., & Sturge-Apple, M. (2012). *Speech-based emotion classification using multiclass SVM with hybrid kernel and thresholding fusion.* Paper presented at the 2012 IEEE Spoken Language Technology Workshop (SLT 2012), 2-5 Dec. 2012, Piscataway, NJ, USA.

Yu, M.-C. (2011). Multi-criteria ABC analysis using artificial-intelligence-based classification techniques. *Expert Systems with Applications, 38*(4), 3416-3421. doi:10.1016/j.eswa.2010.08.127

Zainuddin, Z., & Ong, P. (2011). Reliable multiclass cancer classification of microarray gene expression profiles using an improved wavelet neural network. *Expert Systems with Applications, 38*(11), 13711-13722. doi:10.1016/j.eswa.2011.04.164